# Arabic and contact-induced language change

Edited by

Christopher Lucas
Stefano Manfredi

language
science
press

Series information: ./langsci/seriesinfo/cam-info.tex not found!

# Arabic and contact-induced language change

Edited by

Christopher Lucas

Stefano Manfredi

Freie Universität Berlin

# Contents

*Contents*

# Chapter 1

# Arabic in Iraq, Syria, and Southern Turkey

Stephan Procházka

University of Vienna

This chapter covers the Arabic dialects spoken in the region stretching from the Turkish province of Mersin in the west to Iraq in the east including Lebanon and Syria. The area is characterised by a high degree of linguistic diversity, and for about two and a half millennia Arabic has come in contact with various other Semitic languages as well as with Indo-European languages and Altaic Turkish. Bilingualism, particularly with Aramaic, Kurdish, and Turkish, has resulted in numerous contact-induced changes in all realms of grammar including morphology and syntax.

gll has to be readded to many examples

## 1 Introduction

## 2 Current state and historical development

The region discussed in this chapter is linguistically extremely heterogeneous: in it three different Arabic dialect-groups plus several other languages are spoken. The two main Arabic dialect-groups are Syrian and Iraqi, the distribution of which does not exactly correspond to the political boundaries of those two countries. Syrian-type dialects are also spoken in Lebanon, in three provinces of southern Turkey (Mersin, Adana,[1] Hatay), and in one village on Cyprus. In Iraq, Arabic is mainly spoken in Mesopotamia proper, whereas considerable parts of the

---

[1] The dialects spoken in Mersin and Adana provinces will henceforth referred to as Cilician Arabic.

mountainous parts of the country are Kurdish speaking. Arabic dialects, which are very akin to the Iraqi ones, extend into northeastern Syria and southeastern Anatolia (for the latter see Akkuş in this volume). These two groups are geographically divided by a third dialect-group which came here with an originally (semi)nomadic population from Northern Arabia. Today, this variety preponderates in all villages and most towns between the eastern outskirts of Aleppo and the left bank of the Tigris, and stretching north into the Turkish province of Şanlıurfa.

The total number of native Arabic speakers in the whole region is estimated to be 54 million (see Table 1). The dialects of large urban centers like Beirut, Damascus, Aleppo, Baghdad, and Mosul have become supra-regional prestige varieties that are also used in the media and therefore understood by most inhabitants of the respective countries. The situation is very different in Turkey, where the local Arabic is in sharp decline and public life exclusively dominated by Turkish. Only recently has the position of Arabic in Turkey been socially enhanced by the influx of nearly 3.5 million Syrian refugees[2] fleeing the civil war that started in 2011.

Table 1: Speaker population.

| Country | Speakers |
|---------|----------|
| Syria | 17,000,000 |
| Lebanon | 6,000,000 |
| Iraq | 30,000,000 |
| Turkey | 1,000,000 |

Arabic was spoken in the region long before the advent of Islam (**Donner1981**) but became the socially dominant language in the wake of the Muslim conquests in the 7th century CE. From that time until the end of the 10th century, when Bedouin tribes seized large parts of central and northern Syria, there was probably a continuum of sedentary-type dialects that stretched from Mesopotamia to the northeastern Mediterranean (**Procházka2018**). During the Mongol sack of Iraq in 1258, much of the population was killed or expelled, resulting in far-reaching demographic and linguistic changes as the original sedentary-type dialects could merely hold ground in Baghdad and the larger settlements to its north. Further south they persisted only among the non-Muslim population. Most of today's Iraq was re-populated by people who spoke bedouin-type dialects

---

[2]http://multeciler.org.tr/turkiyedeki-suriyeli-sayisi/ <visited 22 **Feb2018**>.

(mostly coming from the Arabian Peninsula), which over the centuries have heavily influenced the speech of even most large cities (**Holes2007**). Very similar dialects are spoken further south and in the Iranian province of Khuzestan (see Leitner this volume). The foundation of national states after WW1 caused a significant decrease in contact between the different dialect groups and an almost complete isolation of the Arabic dialects spoken in Turkey.

## 3 Contact languages

During its two and a half millennia presence in the region, Arabic has come into contact with many languages, both Semitic and non-Semitic. Those most relevant for the topic will be treated in more detail below (for Syria, see also Barbot, 1961: 175-177). Akkadian was spoken in southern Iraq until about the turn of the eras, i.e. the 1st century AD.[3] Greek was the language of administration in Greater Syria until the Arab conquest (**Magidow2013**) and continued to play a role for Orthodox Christians.[4] During Crusader times Arabic speakers in Syria came into contact with various medieval European languages; and along the Mediterranean coast the so-called Lingua Franca (see Nolan's chapter in this volume) was an important source for the spread of particularly nautical vocabulary for many centuries (Kahane, Kahane, and **Tietze1958**). Since the 19th century locally restricted contacts between Arabic and Armenian and Circassian have existed in parts of Syria and Lebanon.

### 3.1 Aramaic

Aramaic is a NW Semitic language and thus structurally very similar to Arabic. Different varieties of Aramaic were the main language in Syria and Iraq from the middle of the 1st millennium BCE and it can be assumed that some contact with Arabic existed even at that time. From the 1st century CE onwards the southern fringes of the Fertile Crescent became largely Arabic-dominant and there was significant bilingualism with Aramaic, particularly in the towns along the edge of the steppe, such as Petra, Palmyra, Hatra, and al-Hira (**Procházka2018**). Though after the Muslim conquests Arabic eventually became the majority language, it did not oust Aramaic very quickly: the historical sources suggest that Aramaic dominated in the larger towns and the mountainous regions of Syria and

---

[3]For Akkadian lexical influence on Arabic cf. **Holes2002** and **Krebernik2008**.

[4]The enormous influence of Modern Greek on the Arabic spoken in the Kormakiti village of Cyprus will not be discussed here (for a detailed study see Borg, 1985).

Lebanon for a long time. In Iraq, by contrast, the massive influx of Arabs into the cities fostered their quick Arabization while Aramaic continued to be spoken in the countryside (**Magidow2013**: 184; 188). But over the centuries the diverse Aramaic dialects became marginalized and, with very few exceptions, were finally relegated to non-Muslim religious minorities, particularly Christians and Jews, in peripheral regions like Mount Lebanon and the Anti-Lebanon Mountains, where Aramaic was prevalent until the 18th century (**Retsö2006**). Western Aramaic is still spoken in three Syrian villages the best known of which is Maaloula.[5] There also remain speakers of Neo-Aramaic in northern Iraq.[6]

It is hard to establish the degree of bilingualism in the past, but it can be assumed that it was mostly Aramaic L1 speakers who had a command of Arabic and not vice versa. In the present time, nearly all remaining Aramaic speakers in Syria are fluent in Arabic. In Iraq this is mainly true of those living in the plain just north of Mosul (**ArnoldBehnstedt1993**; **Coghill2015**: 86). The influence of different strata of Aramaic on spoken Arabic is a long debated issue, various scholars rating it from considerable to negligible (**Hopkins1995**: 39; **Lentin2018**).

## 3.2 Persian and Kurdish

For many centuries Arabic and the two Western Iranian Languages Persian and Kurdish have influenced each other on different levels. Persian speaking communities existed in medieval Iraq, and economic and cultural contacts between Mesopotamia and Iran have continued to the present (cf. **Gazsi2011**). An important factor of language contact are the holy shrines of the Imams in Kerbela, Najaf, and other Iraqi cities, which have always attracted tens of thousands of Persian speaking Shiites every year. Intensive contacts between Kurdish and Arabic have existed since at least the 10th century, particularly in Northern Iraq, NE Syria, and SE Anatolia (see Akkuş this volume). Until their exodus in the early 1950s, the Arabic speaking Jewish communities which existed in Iraqi Kurdistan usually had a native-like command of Kurdish (**Jastrow1990**). Due to the multi-lingual character of the region, bilingualism of Kurdish and Arabic is still relatively widespread, particularly in urban settings, though with Kurds usually much more fluent in Arabic than the other way around[7]. However, for obvious reasons little linguistic research has been done in Iraq for decades, which makes

---

[5]The village heavily suffered from the Jihadist occupation 2013-2014, but after government troops had took control over the region again, many inhabitants returned and began its reconstruction (cf. the reports collected at http://friendsofmaaloula.de/).

[6]See **Coghill2015** and http://glottolog.org/resource/languoid/id/nort3241.

[7]With significant exceptions in some parts of SE Anatolia: see Akkuş this volume.

it impossible to give up-to-date information about the linguistic situation in ethnically mixed cities like Kirkuk.

## 3.3  Ottoman and Modern Turkish

Contacts between spoken Arabic varieties and various Turkic languages existed from the 9th century onwards. These early contacts, however, left hardly any traces in Arabic except for a handful of loanwords. In the 16th century, the Ottomans established their rule over most Arab lands, including Syria, Lebanon, and Iraq. This domination lasted four hundred years, until WW1. Particularly in the provinces of Aleppo and Mosul, there were a relatively high percentage of Turkish speakers and probably a significant degree of bilingualism[8]. As the language of the ruling elite, Turkish had high prestige and therefore was at least rudimentarily spoken by many inhabitants of those regions, especially urban men. The collapse of the Ottoman Empire put an abrupt end to Turkish-Arabic contacts, which today remain intensive only among the Arabic varieties spoken within the borders of Turkey itself, where most Arabic speakers are fluent in Turkish, the dominant language in all contact settings.

In some areas of Syria and in Northern Iraq the Arabic-speaking population lives side by side with several hundred thousand speakers of Turkish and Azeri Turkish, who call themselves Turkmens. Unfortunately, no reliable data on the socio-linguistic settings and the degree of bilingualism exist for those areas. Again, it can be assumed that most of the Turkmens in both countries are dominant in Turkish, but know Arabic as a second language.

## 3.4  French and English

After WW1 Syria and Lebanon stayed under French mandate and Iraq under British mandate until they reached independence.[9] French is still widely spoken as a second language in Lebanon, especially by Christians. In Iraq, English has maintained its position as by far the most important foreign language – a fact which was reinforced by the US military occupation from 2003 to 2010.

---

[8]Cf. **Wilkins2010**: XV for Aleppo; **Khoury1987**: 103 maintains that Aleppo's hinterland was culturally even more Turkish than Arab. For Mosul see **Shields2004**: 54-55.

[9]Iraq in 1932, Lebanon in 1943, Syria in 1946.

### 3.5 Intra-Arabic contacts

Contacts between different Arabic varieties, for instance between speakers of rural and urban dialects, happen on an everyday basis and often trigger short-term accommodation without leading to long- lasting changes. The situation is different with regard to the enduring contacts between the Bedouin and the sedentary populations, whose dialects considerably differ from each other[10]. Such contacts are most intense at the periphery of the Syrian Steppe and along the Middle Euphrates, where scattered towns with sedentary dialects like Palmyra, Der ez-Zor, and Hit are surrounded by an originally nomadic population. Though the nomadic way of life has been abandoned by most of them they still speak bedouin-type Arabic dialects. As the nomads were, for many centuries, socially and economically dominant, speakers of sedentary dialects often adopted linguistic features from the more prestigious Bedouin (though reverse instances are also found; cf. (Peter **Behnstedt1994b**: 421). Due to the historical circumstances mentioned in §1, Bedouins also had a strong linguistic impact on Iraqi dialects. In Baghdad the sedentary dialect of the Muslim population has been gradually bedouinized due to massive migration from the countryside to the city (**Palva2009**). The Christian and, in former times, Jewish inhabitants preserved their original sedentary-type dialects because they had much less contact with the Muslim newcomers.

## 4 Contact-induced changes

Change induced by contact with Aramaic almost exclusively happened through imposition, i.e. by Aramaic speakers who had learned Arabic as a second language and later often completely shifted to Arabic. This explains the relatively numerous phonological changes and pattern replications in syntax. Lexical transfers from Aramaic certainly were also made by Arabic-dominant speakers, particularly in semantic fields like agriculture that included novel concepts for the mostly animal breeding Arabs.

The same is true for transfers from Greek, for which a very low level of bilingualism can be assumed. Thus we find only matter replication in the form of loanwords, mostly in domains where lexical gaps in older layers of spoken Arabic are likely.

---

[10]Since these two speech communities differ from each other in so many ways it is a relatively robust approach to rate the following features as results of dialect contact and not mere variation (cf. Lucas, 2015: 533).

In the case of Kurdish, bilingualism is much more widespread among speakers of the source language, suggesting imposition. This might explain the phonological changes as speakers dominant in the source language tend to preserve its phonological features (**Lucas2015**). The relatively small number of lexical matter replication is probably the result of the fact that Arabic has long been regarded as the more prestigious by speakers of both the source and the recipient language.

The numerous loanwords from Persian into Iraqi Arabic may well be the result of matter replication by agents who were dominant in the recipient language Arabic. Starting with the rule of the Abbasid caliphs in the 8th century and continuing to the present, Iranian material culture and cuisine often had a great impact on neighboring Mesopotamia. There were also many intellectuals, among them praised writers of Arabic prose, who were actually Iranians and hence knew both languages. Frequent contacts on the everyday level caused additional borrowing of ordinary vocabulary and the retention of sounds that are replaced in Persian loans found in CA or other dialects.[11]

Changes induced by contact with Ottoman Turkish may have happened mostly through Arabic-dominant speakers. The current situation of Arabic speakers in Turkey is, however, very different because at least the last two generations have acquired Turkish as L2 or even as a second L1 at very young age. Thus, at least some of the contact phenomena described in the following paragraphs may be examples of linguistic convergence (see also **Lucas2015**: 525).

French and English have largely remained typical "foreign languages" learned at school or in business with a considerable amount of bilingualism only in some urban settings of Lebanon, particularly Beirut. The agents of change are certainly dominant in the recipient language.

The distinction between the two transfer types is not always clearly discernable in case of intra-Arabic contact-induced changes. In the towns of the Syrian Steppe and the Middle Euphrates the agents of change were mostly the sedentary population who adapted their speech towards the "norms" of the socially more prestigious Bedouin. However, there has always been inter-marriage, and Bedouins often settled in towns and may well have adopted features from the local sedentary variety. Especially in cases like Muslim Baghdadi (see §1.) we may assume with good reasons that the bedouin character of today's variety developed through both imposition and borrowing.

---

[11]The phonological changes have, however, more than only the Persian "background" (cf. §3.1).

## 4.1 Phonology

### 4.1.1 Aramaic-induced changes

It has been hypothesized that several phonological features of the Syrian and Lebanese dialects are due to contact-induced influence of Aramaic. But in the case of the shift from interdental fricatives to postdental plosives (*ḏ > d; ṯ > t; ḏ̣ > ḍ*) this is unlikely because (1) this sound change is common cross-linguistically, (2) it does not occur in all dialects of the region, and (3) it is found in many other Arabic dialects without an Aramaic substrate.

A phonotactic characteristic of most dialects spoken along the Mediterranean, from Cilicia in the north to Beirut in the south, is that all unstressed short vowels (including /a/) in open syllables are elided whereas in other dialects east of Libya only /i/ and /u/ in this position are consistently dropped.[12]

(1)  Cilician Arabic (**Procházka2002b**: 31-32; 130)
     OA raṣāṣ > rṣāṣ 'lead, plumb'
     OA miknasa > mikinsi 'broom'[13]
     fataḥ + t > ftaḥt 'I opened'

Because this rule corresponds to the phonotactics of Aramaic and is otherwise not found in the same degree except in Maghrebinian dialects, pattern replication is likely though cannot be proved. [14]

In roughly the same region except Cilicia and many dialects of Hatay,[15] the diphthongs /ay/ and /aw/ are only preserved in open syllables, but monophthongised to /ē/ and /ō/ respectively in closed syllables. In some regions, for instance on the island of Arwad, both diphthongs merge to /ā/ in closed syllables (**Behnstedt1997**: map 31).

(2)  Arwad, W-Syria (**Procházka2013**)
     OA bayt / baytayn > bāt / baytān 'house – two houses'
     OA yawm / yawmayn > yām / yawmān 'day – two days'
     OA bayn al-iṯnayn > bān it-tnān 'between the two'

Likewise, in older layers of Aramaic, diphthongs were usually monophthongised in closed syllables (for Syriac see Nöldeke, 1904: 34), which makes imposition by L1 speakers of Aramaic rather likely (**Fleisch1974c**).

---

[12]Therefore, **Cantineau1960** called them *parlers non différentiels* – a term still very often applied in Arabic dialectology – as they make no distinction in the treatment of the three short vowels.

[13]With insertion of an epenthetic i to avoid a sequence of three consonants.

[14]Cf. **Diem1979**; Arnold and **Behnstedt1993**; **Weninger2011**.

[15]Where this phenomenon occurs only in Alawi villages (**Arnold1998**).

Another striking phenomenon is the split of historical /ā/ into /ō/ and /ē/ that is found in scattered areas of the Levant, particularly Northern Lebanon around the Syrian port of Tartous, the Qalamūn Mountains, and the exclusively Christian town of Maḥarde on the Orontes River[16]. Because in many varieties of Aramaic, the old Semitic /ā/ is reflected as /ō/, it could be assumed that Aramaic speakers transferred their peculiar pronunciation to Arabic when learning it. **Fleisch1974b** rejected the hypothesis of an Aramaic influence, arguing that the conditioned distribution of the two allophones is merely a further development of the [ɒ]: [æ] split widely attested for Lebanon and parts of W Syria. However, in the Syrian Qalamūn Mountains there are dialects with an unconditioned split (**Behnstedt1992**), and this is precisely the region where the shift from Aramaic to Arabic occurred relatively late, probably after a long phase of bilingualism. In the town of Nabk, for instance, one can infer that the former Aramaic speaking inhabitants would have simply turned every *ā* into *ō* – except those which long before had become *ē* (or *ǟ*) as a result of the so-called conditioned *imāla* (i.e. the tendency of long *ā* to be raised towards *ē* or even *ī* if the word contains an *i/ī*)[17]. Example (3) clearly shows that the distribution of the allophones is not conditioned by the consonantal environment.

(3)  Nabk, Syria (**Gralla2006**)
OA ṭābix > ṭǟbex 'cooking' vs. OA ṭālib > ṭōleb 'student'
OA ḥāmil > ḥǟmel 'pregnant' vs. OA ḥāmiḍ > ḥōmeḍ 'sour'

In these cases Aramaic influence seems plausible. For the region of Tripoli it may be assumed that Aramaic bilinguals from the adjacent mountains used *ō* instead of *ā* when speaking Arabic and thus reinforced the already existing [ɒ]: [æ] split.[18]

### 4.1.2 The "new" phonemes /č/, /g/, and /p/

Consonantal phonemes that are originally alien to Arabic are found in all Arabic dialects spoken in Turkey, Northern Syria, and Iraq. These are the unvoiced affricate /č/, the voiced /g/[19], and the unvoiced /p/, the latter mainly used in

---

[16]For details cf. Behnstedt (1997: map 32). The conditioned shift *ā* > *ō* is also found in and around Tarsus in Turkey (**Procházka2002b**).

[17]Cf. Arnold and **Behnstedt1993**.

[18]For discussion see **Fleisch1974b**; **Fleisch1974c**, **Diem1979**, **Behnstedt1992**, (Arnold and **Behnstedt1993**, **Weninger2011**.

[19]The sound *g* is prevalent in whole Syria and Lebanon but seems to have phonemic status only in the north (**Sabuni1980**). For further examples and discussion see **Ferguson1969**. This "foreign" *g*

Iraq. These sounds were very likely contact-induced, but it is often impossible to discern which language triggered each development: all three sounds are found in Persian, Kurdish, Turkish, and the Lingua Franca. For the dialects of Cilicia, Hatay, and Syria the main source language doubtless was Turkish. The sound /p/ in the Iraqi dialects was probably first introduced through contact with Persian and Kurdish, and then reinforced from Ottoman Turkish. In the bedouin-type dialects of the region the phonemes /č/ and /g/ are not products of contact-induced change but occur due to internal sound changes, unvoiced /č/ as a conditioned affricated variant of /k/ and /g/ as the ordinary reflex of OA /q/.

Thus, it can be assumed that over the centuries speakers of the sedentary dialects of Iraq and Syria borrowed either from other languages or from bedouin Arabic varieties words that possess these two sounds, which subsequently were fully incorporated into the phonemic inventory. This development may have been facilitated by the fact that the three sounds /č/, /p/, and /g/ are not fundamentally unfamiliar to Arabic but are the voiceless, respectively voiced counterparts of the well-established phonemes /ǧ/, /b/, and /k/. It seems no accident that the new sound /č/ is much more often found in dialects that have preserved the affricate /ǧ/ than in those where it has shifted to /ž/.

(4) Aleppo
  *čanṭāye* 'handbag' (Turkish *çanta*)
  *čwāl* 'sack' (Turkish *çuval*)
  *čāy* 'tea' (Turkish *çay*)
  *gaǧaleg* 'nightgown' (Turkish *gecelik*)

The words given above are usually pronounced with *š* instead of *č* in the central Syrian and Lebanese dialects where contact with Turkish was less intense and /ǧ/ is reflected as /ž/.[20]

(5) Mosul
  *ṣūč* 'fault' (Turkish *suç*)
  *pāča* 'stew of sheep and cow legs and innards' (Kurdish/Persian *pāče*)
  *zangīn* 'rich' (Turkish *zengin*)

Once integrated into the phonological system, these sounds not only enabled easier integration of loanwords from other languages like French and English

---

must therefore be differentiated from *g,* which is the regular reflex of OA *q.* The latter development is found in many bedouin-type dialects.

[20] **Behnstedt1997**: maps 18, 19, 25. For details and more examples cf. **Sabuni1980** (lists all words with *č/g* in Aleppo) and **Procházka2002a** for Cilician Arabic.

(see §3.4), but sometimes also resulted in the spread of assimilation-induced allophones from single words to the whole paradigm or even root. In Aleppo one finds \*yəkdeb > yəgdeb 'he lies' due to assimilation. The *g* subsequently was transferred to other words derived from the root: *gadab* 'he lied', *gədbe* 'lie', and *gaddāb* 'liar' (**Sabuni1980**: 26; 209).

Speakers of sedentary dialects who had everyday contact with Bedouins – for example the inhabitants of Der ez-Zor and Khatuniyya – first integrated /č/ and /g/ into their phonemic inventory through the borrowing of typically Bedouin vocabulary such as *dabča* 'a Bedouin dance' (Khawetna) and *ṭabga* 'milk-bowl' (Soukhne). These sounds then entered other fields of the lexicon, which led to unpredictable distribution, including doublets.

- Khawetna (**Talay1999**)

- *gəṣṣa* 'forehead', but *qəṣṣa* 'story' (OA *quṣṣa / qiṣṣa*)

- *dīč* 'rooster' (OA *dīk*)

- Der iz-Zor (**Jastrow1978**).

- *gāˤ* 'soil' (OA *qāˤ*)

- *čam* 'how much?' (OA *kam*)

- Baghdad (**Palva2009**)

- *guffa* 'large basket' (OA *quffa*), but *quful* 'lock' (OA *qufl*)

- *ˤigab* 'to pass', but *ˤiqab* 'to follow' (both OA *ˤaqab*)

- The opposition /k/: /č/ has even entered morphology, particularly with the 2nd person sɢ suffixes: *ʔabū-k* vs. *ʔabū-č* 'your (M/F) father'.

In the Syrian oasis of Soukhne, long-term contact with speakers of bedouin dialects caused a chain of phonetic changes: first /k/ shifted to /č/, which originally was the reflex of OA /ǧ/; then /č/ (< /ǧ/) shifted further to /tˢ/, which has become a unique feature of the local dialect. The unconditioned shift from /k/ > /č/, which is not found in the bedouin dialects, in turn caused a shift from /q/ > /k/.[21]

---

[21]Cf. (Peter **Behnstedt1994a**: 4-11) for details.

(6)    Soukhne (Peter **Behnstedt1994a**: 226; 344; 357; 360)
       *kirbi* 'water-skin' (< OA *qirba*; bedouin *girba*)
       *čalb* 'dog' (< OA *kalb,* bedouin *čalib*)
       *čurr* 'donkey foal' (< OA *kurr,* bedouin *kuṛṛ*)
       *tˢubn* 'cheese' (< OA *ǧubn,* bedouin *ǧubun*)

## 4.2  Morphology

The Aramaic diminutive suffix *-ūn* has become restrictedly productive in Iraqi Arabic (**Masliyah1997**); in Syria/Lebanon it is only found in petrified forms such as *šalfūn* 'young cock' and *qafṣūne* 'little cage'. Such kind of morphological transfer is usually triggered by lexical borrowing. Thus, it may be assumed that this suffix spread from loanwords like *šalfūne* 'small knife blade' < Aramaic *šelpūnā* 'little knife' (cf. **Féghali1918**: 82).[22] .

(7)    Iraq
       *darb* 'road' > *darbūna* 'alley'
       *gṣayyir* 'short' < *gṣayyrūn* 'very short'
       *mḥammdūn* hypocoristic form of the name Muḥammad

Syrian and Lebanese dialects exhibit a few word patterns that are attested for OA (and other dialects) but seem to have become widespread through contact with Aramaic due to their frequency in the latter: they are the verbal pattern šaC$_1$C$_2$aC$_3$ and the (primarily diminutive) nominal patterns C$_1$aC$_2$C$_2$ūC$_3$ and C$_1$aC$_2$C$_3$ūC$_4$.[23] An example for the first is *šanfax* 'to puff up', related to *nafax* 'to blow up' (**Féghali1918**); for discussion cf. **Lentin2018**.

(8)    Aleppo (**Barthélemy1935**: 104; 158; 851)
       *ǧaḥḥūš* 'little donkey'; related to *ǧaḥš*
       *ḥassūn* 'goldfinch'; the small bird's name is related to *ḥasan.*
       *namnūme* 'small louse'

The pattern C$_1$aC$_2$C$_2$ūC$_3$(i) is still productive in the whole region including the bedouin dialects to derive hypocoristic forms from personal names:

---

[22]This must be a very old borrowing because the suffix is also found in the Gulf dialects (e.g. *ḥabbūna* 'a little', **Holes2002** and even in Tunisian Arabic (**Singer1984**) where direct Aramaic influence can be excluded.

[23]For the latter two cf. **Corriente1969** and **Procházka2004**.

(9)  Fāṭma > Faṭṭūma
     Ḥalīme > Ḥallūma
     Aḥmad/Mḥammad > Ḥammūdi

In all Syrian and Lebanese dialects the pronouns of the 2nd and 3rd persons PL exhibit an *n* in place of the *m* that is found in other Arabic dialects, which makes them look as if they were reflexes of OA feminine forms (Table 2).

Table 2: Personal pronouns 2

|       | Damascus | Jerusalem | OA F.PL | Syriac M.PL |
|-------|----------|-----------|---------|-------------|
| 2PL   | *ʾəntu / -kon* | *ʾintu / -kom* | *ʾantunna / -kunna* | *ʾatton / -kon* |
| 3PL   | *hənne(n) / hon* | *humme / -hom* | *hunna / -hunna* | *hennon / -hon* |

[nd] and 3rd persons plural

Because generalization of the feminine is unlikely[24] , these forms have often been explained as a contact-induced change. In Aramaic the corresponding pronouns also have *n* (for Syriac cf. Muraoka, 2005: 18). In particular, the 3rd person forms with final -*n* exactly mirror the Aramaic pattern, but lack a plausible intra-Arabic etymology. Thus imposition seems plausible. Nevertheless, substratum influence has been doubted, particularly because of the infrequent evidence of *n*-pronouns in other regions.[25]

Also, the vowels -*o* in the west of the region and -*u* in the east can be affixed to various kinship terms and given names when used for direct address, usually hypocoristically.[26]

(10)  Urfa (own data)
      *šnōnak xayy-o?* 'Brother, how are you?'
      *ǧidd-o* 'Grandfather!'
      *ʿamm-o* '(paternal) Uncle!'
      *xāl-o* '(maternal) Uncle!'

---

[24]Mainly, because the feminine forms are only used for addressing a group of females, whereas the masculine forms may also refer to a mixed group. Therefore, the masculine forms are certainly more frequent. In all Arabic dialects except those mentioned above the gender neutral plural forms are clearly derived from the historical masculine.

[25]Cf. **Owens2006** and **Procházka2018** for details.

[26]See also **Ferguson1997**.

In Syria the suffix is also added to female nouns, i.e. ˁ*amm-t-o* '(paternal) Aunt!' and *xāl-t-o* '(maternal) Aunt!', whereas in Iraq the corresponding forms end in *-a*: ˁ*amm-a, xāl-a.*

Since this suffix has no overt Arabic etymology[27] it has been assumed to be a borrowing of the Kurdish vocative *-o* (e.g. Grigore, 2007: 203). The Persian suffix *-u* also forms affective diminutives[28], which would make Persian influence possible, at least for Iraq.[29] However, the distribution of this feature is far beyond even indirect contact with Kurdish or Persian[30], though reinforcement and influence on the phonology may be possible for certain regions. Similar endings in Aramaic (**Fassberg2010**) and Ethiopian (**Brockelmann1928**) suggest a common Semitic origin (see also Pat-El, 2017: 463-465).

All dialects of the region have incorporated the Turkish suffix /-ci/ [ʤi] into their nominal morphology. This suffix has become productive and is therefore a good example of morphological matter-borrowing (Gardani, Arkadiev, und Amiridze, 2015). It is widely used for expressing professions, occupations, and habitual actions – the latter overwhelmingly pejorative, or at least humorous. In Iraqi dialects the suffix is reflected as *-či*. In the other varieties, it follows the usual development of /ǧ/, which means that it is pronounced *-ǧi* or *-ži*.

(11)  Syria
    *kahrab-ži* 'electrician' (*kahraba* 'electricity')
    *nəswān-ži* 'womanizer' (*nəswān* 'women')
    *maškal-ži* 'troublemaker' (*məšᵊkle* 'problem')

(12)  Iraq
    *pančar-či* 'tire repairman' (*pančar* 'puncture')
    *mharrib-či* 'human trafficker' (*mharrib* 'one who helps so. to escape')
    ˁ*arag-či* 'drunkard' (ˁ*arag* 'aniseed brandy')

The suffix clearly fills a morphological gap because it enables morphologically transparent derivation even from loanwords by preserving the basic, immedi-

---

[27]I warmly thank Jérôme Lentin for extensive discussion of this issue and invaluable help in finding important sources.

[28]E.g. *pesar-u* 'kid'; ˁ*amm-u* is even the common word for 'uncle' (**Perry2007**).

[29]In the Iraqi dialects the vowel is *-u*, e.g. ˁ*amm-u, xāl-u* and *ǧidd-u* (Abu-**Haidar1999**: 145).

[30]The suffix is, for instance, attached to given names for endearment in the Gulf dialects, cf. **Holes2016**: 128. The address forms *ya* ˁ*amm-u, ya xāl-u* 'uncle', *gidd-u* 'grandfather', *sitt-u* 'grandmother' are used in Cairo where hypocoristic variants of given names are likewise attested, e.g. *Mīšu* for Hišām (**Woidich2006**). The suffix *-o/-u* in address forms is also attested in eastern Sudan (Stefano Manfredi, pc.) and in the Maghreb (**PrunetIdrissi2014**: 184 provide a list of such nouns for Morocco).

ately recognizable, word – in contrast to the Arabic $C_1aC_2C_2\bar{a}C_3$-pattern or participles, which are derived from the root (for details cf. Procházka-Eisl, 2018).

To a lesser extent other Turkish suffixes have enhanced the morphological devices of the dialects treated here [31], specifically the relative suffix /-li/, the privative suffix /-siz/, and the abstract suffix /-lik/, which is reflected as *-loġiyya* in Iraq, i.e. with the Arabic abstract morpheme affixed. For the most part these suffixes appear in Turkish loanwords, e.g. Cilicia *ṣiḥḥat-li* (< Turkish *sıhhatlı*) 'healthy', *raḥaṭ-ṣīz* (< Turkish *rahatsız*) 'uncomfortable'. Only in Iraq have they gained a certain degree of productivity, particularly *-sizz* and *-loġiyya*.

(13)   Iraq (**Masliyah1996**)
       *muxx-sizz* 'stupid, brainless'
       *ḥaya-sizz* 'shameless'
       *ḥaywān-loġiyya* 'ignorance' (lit. 'animal-ness')
       *zmāl-loġiyya* 'stupidity' (lit. 'donkey-ness')

Arabic dialects spoken in Turkey not infrequently use light verb constructions (in Turkish grammar mostly called phrasal verbs) which consist of the verb 'to do' plus a following noun. Such compound verbs are very frequent in Turkish (and Kurdish) and enable easy integration of foreign vocabulary into the verbal system. The light verbs found in the Arabic dialects show that this formation is a case of selected pattern replication because, first, not all examples are exact copies of the Turkish model, and second, the word-order follows the Arabic V-O rather than the Turkish O-V pattern.

(14)   Harran-Urfa (own data)
       *sāwa qaza* (Turkish *kaza yapmak*) 'to have an accident'
       *sāwa ʕēš* (in Turkish not a phrasal verb, but *pişirmek*) 'to cook'

(15)   Cilician Arabic (**Procházka2002b**)
       *sawwa zarar* (Turkish *zarar vermek*) 'to harm'
       *sawwa xayir* (Turkish *hayır işlemek*) 'to do a good deed'

Intra-Arabic contact led to the adoption of typical bedouin-type pronouns into sedentary dialects (cf. **Palva2009**: 27-29).

(16)   Baghdad, Der ez-Zor, Soukhne
       *ʔəḥna* for *nəḥna* 'we'

---

[31]Cf. Halasi-**Kun1969**, **Sabuni1980**, **Masliyah1996**, **Procházka2002b**.

(17)  Baghdad
      *ʾāni* for *ʾana* 'I'

As shown in Table 3, virtually all the eastern sedentary dialects of Syria have copied the typical bedouin-type active participles of the verbs 'to eat' and 'to take' that exhibit initial *m-*. (**Behnstedt1997**: map 175).

Table 3: Active participles of the verbs 'to eat' / 'to take'

| **Bedouin** | **Soukhne** | **Palmyra** | **Damascus** |
|---|---|---|---|
| *māčil / māxiḏ* | *mīčil / mīxiḏ* | *mākil / māxiḏ* | *ʾākel / ʾāxed* |

In a few places intensive mutual contact has resulted in an interdialect (**Trudgill1986**) with completely new forms (Table 4), such as the inflectional suffix *-a* in the Syrian village of Ṣōrān (Peter **Behnstedt1994c**: 423-425)

Table 4: Inflectional suffixes 3

| **Bedouin** | **Sedentary** | **Ṣōrān** |
|---|---|---|
| *gāḷ-am* | *qāl-o* | *qāl-a* |
| | M.PL | |

## 4.3  Syntax

In all but the bedouin-type dialects of the region two constructions exist which both use an anticipatory pronoun and the preposition *l-* 'to' for (1) analytical marking of a definite direct object (Examples 4-6) and (2) analytic attribution of a noun (Example 7). The frequency and constraints of these two cases of clitic doubling show great variety but in general the usage of construction (1) is restricted to specific objects, particularly elements denoting human beings, and construction (2) is mostly found with inalienable possessions, particularly kinship. A detailed discussion of both features is found in **Souag2017**.

(18)  Damascus (**Berlinches2016**)
      ḥabb-ēt-o la-ˁamᵊr
      love.PRF-1SG-3M.SG to-Amr
      'I loved Amr.'

(19)   Baghdad, Christian (Abu-**Haidar1991**: 116)
       qaġ-ētū-nu l-əl-əktēb
       read.PRF-1SG-3M.SG to-DEF-book
       'I read the book.'

(20)   Cilician Arabic (ʿalā instead of *l-*; (**Procházka2002b**)
       bi-yḥibb-u ʿala xāl-u
       IND-love.IPF.3M.SG-3M.SG on uncle-3M.SG
       'He loves his (maternal) uncle.'

(21)   Baghdad, Christian (Abu-**Haidar1991**: 116)
       maġ-t-u l-axū-yi
       wife-F-3M.SG to-brother-1SG
       'my brother's wife'

Though the preposition *l-* is sometimes attested in CA for introducing direct objects and even in MSA is common for analytic noun annexation, there are good arguments that the two constructions[32] are pattern replications of an Aramaic model[33]. For one thing, they do not have direct parallels either in OA or in dialects which lacked contact with Aramaic. Examples (8-9) show that both constructions have striking parallels in especially the later eastern varieties of Aramaic (**Rubin2005**).

(22)   a.   Syriac (**Rubin2005**)
            bnā-y l-bayt-ā
            build.PRF.3M.SG-3M.SG to-house-DEF
       b.   Syriac (**Hopkins1997**)[34]
            šm-ēh l-gabr-ā
            name-3M.SG to-man-DEF
            'the name of the man'

In the entire western part of the region including S Turkey the preposition *fī* 'in' together with a pronominal suffix is used to express a capability. This has a striking parallel in the modern Aramaic *ʾīṯ b-* 'there is in ~ be able' (**Borg2004**).

---

[32]Not discussed here are two variants of construction (A), one without suffix and the other without preposition (cf. Lentin, 2018).

[33]Among the many studies that are in favor of Aramaic influence are **Contini1999**, **Blanc1964**, **Weninger2011**. **Diem1979** and **Lentin2018** are more skeptical. **Souag2017** suggests that at least "the initial stages of the development of clitic doubling in the Levant derive from Aramaic substratum influence, but the current situation also reflects subsequent Arabic-internal developments."

[34]The same pattern using the linker *d-* is more common.

(23)  Damascus (**Cowell1964**)
      fī-ni sāˁd-ak ᵊb-kamm lēra
      in-1SG help.IPF.1SG-2M.SG with-some pound.SG
      'Can I help you with a few pounds?'

A final example of possible Aramaic influence is the Syrian particle (or prefix?) *šī* that mainly indicates partial specifity. It might be a pattern replication of the Neo-West-Aramaic *mett*, used with the same function (**Diem1979**). What reduces the likelihood of imposition by Aramaic speakers is the existence of a cognate in Moroccan Arabic which is used with almost the same function.[35]

(24)  Damascus (own data)
      hnīk fī šī ʕāmūd
      there there.is IDF column.SG
      'There is some column.'

In the dialects of the Jews of Kurdistan the definite article is often omitted in subject position – a flagrant imitation of the Kurdish model.

(25)  Kurdistan Arabic (**Jastrow1990**)
      baˁdēn mudīr-a baˁat-ət xalf-na
      then director.IDF-F send.PRF-3F.SG after-1PL
      'Then the director let us come.'

An interesting case of calquing which shows the difficulty to distinguish borrowing from imposition (see Manfredi's chapter in this volume) is the conjunction *m-bōr* 'because, in order to'. It exhibits both matter and pattern transfer as it is a copy of Kurdish *ji ber (ku)*. In the actual form the Kurdish *ji* 'from' was replaced by the Arabic equivalent *m-* (**Jastrow1979**).

Syntactic change because of contact with Turkish is restricted to the Arabic dialects spoken in Turkey. In Cilicia and the Harran-Urfa active participles express evidentiality, i.e. they are used in utterances where a speaker refers to second-hand information. As evidentiality is not a common category in Semitic, it is very likely that the bilingual Arabic speakers of those regions copied this linguistic category from Turkish. In Turkish, any second-hand information is obligatorily marked by the verbal suffix /-mIş/ whose second function besides evidentiality is to express stativity and perfectivity. The latter two functions are assumed by the active participle in many Arabic dialects including those in question. Thus,

---

[35]Cf. Brustad (2000: 19; 26-27); **Wilmsen2014**.

we can suppose that the stative/perfective function, which is commonly shared by both Arabic active participles and the Turkish suffix /-mIş/, was likely the starting point of the development that led to the additional evidential function of Arabic participles. The fact that evidentials seem to spread readily through language contact (**Aikhenvald2004**) makes Turkish influence even more probable.[36] Example (13) illustrates how the speaker uses perfect forms for those parts of the narrative he witnessed himself, and participles for secondhand information (perfect forms non-italic, participles in bold face).

(26)  Harran-Urfa (**ProcházkaBatan2016**: 465)
       ʾiḥne b-zimānāt čān ʿid-na ǧār b-al-maḥalle.
       huwwa māt ᵊrtiḥam ᵊngūl-lu Šēx Məṭar [...]
       nahāṛ rabīʿ-u wāḥad **ʿāzm**-u ʿala Stanbūl.
       **rāyiḥ** maʿzūm ʿala Stanbul **māxiḏ** Šēx Məṭar ᵊb-sāgt-u.
       'Once we had a neighbor in our quarter. He died; he passed away. We called him Sheikh Məṭar. One day somebody invited his friend to Istanbul. As he was invited he went to Istanbul and he took Sheikh Məṭar with him.'

In most Arabic dialects that are spoken in Turkey, comparatives and superlatives may be expressed by means of the Turkish particles *daha* and *en*, respectively, followed by the simplex instead of the elative form of the adjective. As for comparatives, the use of such constructions is rather restricted while, at least in Cilician Arabic, they are relatively frequent for the superlative.

(27)  Harran-Urfa (own data)
       daha zēn ṣār-at
       more good become.PRF-3F.SG
       'It has become better.'

(28)  Cilician Arabic (**Procházka2002b**)
       mīn en zangīl bi-d-dini
       who SUPERLATIVE.PARTICLE rich in-DEF-world
       'Who is the richest (person) in the world?'

In Cilicia, comparison is often expressed by the elative pattern of an adjective, which is preceded by the particle *issa*. This clearly reflects a calque: the Turkish

---

[36]For more examples and further details cf. (**Procházka2002b**) **Procházka2002b**: 200–201 for Cilicia and (Procházka and **Batan2016** for the bedouin-type dialects in the Harran-Urfa region.

equivalent of the adverb *issa* 'still, yet' is *daha*, which in Turkish is also used as the particle of the comparative.

(29) Cilician Arabic (**Procházka2002b**)
ṣāyir issa aḥsan
become.PTCP more good.ELATIVE
'It became better'.

(30) Turkish
daha iyi ol-du.
more good become-PRF.3SG
'It became better.'

Sometimes a change in verb valency occurs as a consequence of the copying of Turkish models. A case found throughout these dialects is the verb /ˀaǧab/ 'to like': usually in Arabic the object of liking is the grammatical subject and the person who likes something the direct object of the verb; but in the Arabic dialects in question, the construction of this verb reflects its Turkish (and English) usage with the person doing the liking being the grammatical subject.

(31) Arabic

a. Cilicia (**Procházka2002b**)
ˀǧab-t bayt-ak
like.PRF-1SG house-2M.SG

b. Damascus (own data)
bēt-ak ˀažab-ni
house-2M.SG like.PRF.3M.SG-1SG
'I liked your house.'

## 4.4 Lexicon

Apart from the Aramaic loanwords also found in CA (see **Retsö2006** and Van Putten's chapter in this volume) – often in the realms of religion and cult – the dialects of this region exhibit a large number of Aramaic lexemes. They are particularly common in Lebanon and W Syria, but also found in Iraq and even in the bedouin-type dialects (**Féghali1918**), **Borg2008**, 2004). A large percentage of these words belong to flora and fauna, agriculture, architecture, tools, kitchen utensils, and other material things.[37]

---

[37]Cf. also **Neishtadt2015**.

(32)  *ṣumd ~ ṣimd* 'plough' < Syriac *ṣāmdē* 'yoke'
      *qālūz* 'bolt (of a door)' < Syriac *qālūzā*
      *nāṭūr* 'guard (of a vineyard etc.)' < Syriac *nāṭūrā*
      *šaṭaḥ* 'to spread' < Syriac *šᵉṭaḥ*
      *šōb* 'heat, hot' < Syriac *šawbā*

Many nautical terms and words denoting agricultural products and tools were borrowed by Arabic from Greek, often via other languages, particularly Aramaic[38], the Lingua Franca, and Turkish.

(33)  *brāṣa* < Greek *práson* 'leek'
      *laxana* < Greek *láxana* 'cabbage'
      *dərrāˀen* < Greek *dōrákinon* 'peaches'
      *ˀabrīm/brīm* 'keel' < Greek *prýmnē* 'stern, poop'
      *sfīn* < Greek *sphēn* 'wedge'

Kurdish borrowings are mainly restricted to N Iraq where bilingualism is widespread.

(34)  Mosul
      *pūš* 'chaff' < Kurdish *pûş*
      *hēdi hēdi* 'slowly' < Kurdish *hêdî* (**Jastrow1979**)

The intensive cultural and economic contacts between Iraq and Iran led to many Persian loanwords in various domains of the Iraqi dialects.

(35)  *mēwa* 'fruit' < Persian *mīva ~ mayva*
      *baxat* 'luck' < Persian *baxt*
      *čarix* 'wheel' < Persian *čarx*
      *gulguli* 'pink' < Persian *gol* 'rose'
      *yawāš* 'slow' < Persian *yavāš*
      *puxta* 'mush' < Persian *poxte* '(well) cooked'

Ottoman Turkish contributed a great deal to the culinary vocabulary and the terminology of clothing and (technical) tools of Syria and Iraq[39] . It also borrowed several adverbs and even verbs into the local Arabic (Halasi-**Kun1969**, 1973, 1982).

---

[38]This is especially true for words related to Christian liturgy and ritual, which constitute about 20 per cent of the Greek vocabulary that entered the dialects of Syria.

[39]The same loanwords are, of course, often found in other regions that were under Ottoman rule, above all in Egypt, but also in Tunisia, Yemen and other regions.

(36)    Syria (Damascus)

  *šāwərma* 'shawurma' < Turkish *çevirme*

  *ṣāž* 'iron plate for making bread' < Turkish *saç*

  *yalanži* 'vine-leaves stuffed with rice' < Turkish *yalancı* 'liar' (as they pretend to be "real" *dolma* stuffed with meat)

  *šīš ṭāwūʾ* 'spit-roasted chicken' < Turkish *şiş tavuk*

  *kəzlok* 'glasses' < Turkish *gözlük*

  *ʾūḍa* 'room' < Turkish *oda*

  *ballaš* 'to begin' < Turkish *başla-mak* by metathesis.

(37)    Iraq (Muslim Baghdadi, cf. (**Reinkowski1995**))

  *qūzi* 'a dish with roasted mutton' < Turkish *kuzu* 'lamb'

  *tēl* 'wire' < Turkish *tel*

  *yašmāġ* 'kerchief (for men)' < Turkish *yaşmak* 'veil (for women)'

  *bōš* 'empty; neutral' which yielded also the verb *bawwaš* 'to put into neutral (gear)' < Turkish *boş* 'empty'

  *qačaġ* 'smuggled goods' < Turkish *kaçak*

During the last century the Arabic dialects in Turkey[40] incorporated numerous Turkish words in addition to loanwords from Ottoman times. Among them are terms in education, medicine, sports, media, and technology. Besides these, kinship terms, the vocabulary of everyday life, and structural words like adverbs and discourse markers have infiltrated the dialects from Turkish.

(38)    Cilician Arabic

  *qāyin* ... '-in-law' (< Turkish *kayın*)

  *ṭōrūn* 'grandchild' (< Turkish *torun*)

  *bīle* 'even' (< Turkish *bile*)

  *qāršīt* 'opposite from' (< Turkish *karşı*)

The cases of semantic extension of an Arabic word result from the wider semantic range of its Turkish equivalent which has been transferred into Arabic. Thus, in both Cilician and Harran-Urfa Arabic *sāq/ysūq* 'to drive' also occurs with the meaning of 'to last' like the Turkish verb *sürmek*. In Harran-Urfa *b-arḍ* 'on the place/ground (of)' has become a preposition/conjunction meaning 'instead'. This can be seen as a contact-induced grammaticalization (Gardani, Arkadiev, and **Amiridze2015**: 4) under the influence of Turkish *yerine* 'instead; to its place'.

---

[40]For Cilicia cf. (**Procházka2002b**) and (**Procházka2002a**).

(39)   Harran-Urfa (own data)
       al-mille tākl-u b-arḍ al-laḥam
       DEF-people eat.IPF.3F.SG-3M.SG in-place DEF-meat
       'The people eat it instead of meat.'

(40)   Harran-Urfa (own data)
       b-arḍ-in tibči ʾigir āya
       in-place-LINKER cry.IPF.2M.SG read.IMP.M.SG verse
       'Instead of crying recite a (Koranic) verse!'

In Iraq, many English words related to Western culture and technology have been, and still are being, borrowed into the dialects. The same is true for French in Syria and (particularly) Lebanon (cf. **Barbot1961**: 176).

(41)   Iraq (words of English origin)
       *kitli* < kettle
       *buṭil* < bottle
       *glāṣ* < glass
       *pančar* 'flat tire' (< *puncture*)
       *pāysikil* < bicycle
       *māṭōrsikil* < motorcycle
       *lōri* < lorry
       *igzōz* < exhaust (pipe)
       *brēk* < brake

(42)   Syria and Lebanon (words of French origin)
       *gātto ~ gaṭō < gâteau* 'cake'
       *garsōn < garçon* 'waiter'
       *sēšwār < séchoir* 'hair drier'
       *kwaffēr < coiffeur* 'hair-dresser'
       *ʾaṣanṣēr < ascenseur* 'elevator'
       *grīb < grippe* 'influenza'

Due to long-term contacts, there are mutual borrowings between the bedouin and sedentary dialects of the region. This affects not only specific vocabulary of respective culture but also basic lexical items. Historically, the sedentary dialects have been much more influenced by the bedouin-type dialects than vice versa.

## 5 Conclusion

The sociolinguistic history of the regions treated here suggests that the conditions for imposition were relatively restrictive and mainly found in contact settings with Aramaic which, over the centuries, has been given up by most of its speakers in favor of Arabic. Thus, it is not surprising that so many features beyond the lexicon for which contact-induced change can be assumed are related to Aramaic influence.

Morphological borrowing is in general relatively rare because it presupposes a high intensity of contact (Gardani, Arkadiev, and **Amiridze2015**: 1). Practically all cases presented in §3.2 corroborate the universal tendencies that (1) derivational morphology is more prone to borrowing than inflectional morphology and (2) nominalizers and diminutives are very frequently represented in instances of borrowed derivational morphology (Gardani, Arkadiev, und Amiridze, 2015: 7); **Seifart2013**). On the whole, the bedouin-type dialects exhibit significantly less contact-induced changes than the sedentary dialects. This may be the result of both the production systems of Bedouin groups and their tribally organized society, which impedes intense contact with outsiders.

The relative infrequency of contact-induced changes in morphology and syntax found in the Arabic varieties spoken in Turkey have two main reasons: First, the high degree of complete bilingualism is a very recent phenomenon that only pertains to the last two generations. Second, and probably more important, the great structural differences between the two languages, which have impeded both matter- and pattern-replications.

What is still relatively unclear is the degree of historical bilingualism between Arabic on the one hand and Ottoman Turkish, Kurdish, and Persian on the other. Future research would be particularly desirable with regard to Iraq, providing interesting new data on contact-induced changes in multilingual regions like Mosul and Kirkuk where Arabic, Turkmen, and Kurdish speakers have been in contact for a long time. Also, studies like (**Neishtadt2015**) for Palestine should be carried out for Syrian and especially Iraqi dialects with regard to lexical borrowings from Aramaic. Another completely under-researched topic is idiomatic constructions, in which the mutual influence of most languages in the region may be assumed.

## Further reading

There are no studies which treat the subject with regard to the region as a whole and contacts between Arabic and different other languages.

**(ArnoldBehnstedt1993)**

In-depth study on the mutual contacts between Neo-Western Aramaic and the local Arabic dialects in the Anti-Lebanon Mountains of Syria.

**(Diem1979)**

Pioneer study on substrate influence in the modern Arabic dialects, though with focus on South Arabia, i.e. outside of the region treated in this chapter.

**(Palva2009)**

Very good case study on the diachronic relations between sedentary and bedouin-type dialects in the Iraqi capital Baghdad.

**(Weninger2011)**

Concise overview of the contact between different varieties of Aramaic and Arabic.

## Abbreviations

OA Old Arabic
  CA Classical Arabic
  MSA Modern Standard Arabic

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}

\begin{styleBibliography}
@@book{Abu-Haidar1991,
address = {Wiesbaden},
author = {Abu-Haidar, Farida.},
publisher = {Harrassowitz},
title = {\textit{{Christian}-Arabic} \textit{of} \textit{Baghdad}},
year = {1991}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{———1999,
author = {———},
journal = {„Vocatives as exclamatory nouns in Iraqi Arabic" Richmond: Curzon},
pages = {144–60},
volume = {1999},
```

```
year = {1999}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Aikhenvald2004,
address = {Oxford - New York},
author = {Aikhenvald, Alexandra Y.},
publisher = {Oxford University Press},
title = {\textit{Evidentiality}},
year = {2004}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Arnold1998,
address = {Wiesbaden},
author = {Arnold, Werner.},
publisher = {Harrassowitz},
title = {\textit{Die} \textit{arabischen} \textit{{Dialekte}} \textit{Antiochien
year = {1998}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Arnold1993,
address = {Wiesbaden},
author = {Arnold, Werner, und Peter Behnstedt.},
publisher = {Harrassowitz},
title = {\textit{Arabisch-aramäische} \textit{{Sprachbeziehungen}} \textit{im} \
year = {1993}
}
\end{styleBibliography}


\begin{styleBibliography}
@@article{Barbot1961,
author = {Barbot, Michel},
journal = {\textit{Arabica}},
```

```
number = {2},
pages = {174–88},
title = {„Emprunts et phonologie dans les dialectes citadins syro-libanais"},
volume = {8},
year = {1961}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Barthélemy1935,
address = {Paris},
author = {Barthélemy, A.},
publisher = {Paul Geuthner},
title = {\textit{Dictionnaire} \textit{Arabe-Français,} \textit{dialectes} \textit
year = {1935}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Behnstedt1992,
author = {Behnstedt, Peter},
journal = {\textit{Die} \textit{Welt} \textit{des} \textit{Orients}},
pages = {77–100},
title = {„Zur Spaltung von a in ō/ē (ọ̄/\={ẹ}) im Qalamūn (Antilibanon)"},
volume = {23},
year = {1992}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{———.1994,
address = {Wiesbaden},
author = {———.},
publisher = {Harrassowitz},
title = {\textit{Der} \textit{arabische} \textit{{Dialekt}} \textit{von} \textit{S
year = {1994a}
}
\end{styleBibliography}
```

\begin{styleBibliography}
———. 1994b. „Dialektkontakt in Syrien". In \textit{Actes} \textit{des} \textit{p}
430. Paris: INALCO, Publications Langues'O.
\end{styleBibliography}

\begin{styleBibliography}
@@book{———.1997,
address = {Wiesbaden},
author = {———.},
publisher = {Harrassowitz},
title = {\textit{{Sprachatlas}} \textit{von} \textit{Syrien.} \textit{I:} \texti
year = {1997}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Berlinches2016,
address = {Zaragoza},
author = {Berlinches, Carmen.},
publisher = {Prensas de la Universidad de Zaragoza},
title = {\textit{El} \textit{dialecto} \textit{árabe} \textit{de} \textit{Damasc
year = {2016}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Blanc1964,
address = {Cambridge, Mass.},
author = {Blanc, Haim.},
publisher = {Center for Middle Eastern Studies of Harvard University},
title = {\textit{Communal} \textit{Dialects} \textit{in} \textit{Baghdad}},
year = {1964}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Borg1985,

```
address = {Stuttgart},
author = {Borg, Alexander.},
publisher = {Steiner},
title = {\textit{Cypriot} \textit{{Arabic}}},
year = {1985}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{———.2004,
address = {Leiden},
author = {———.},
publisher = {Brill},
title = {\textit{A} \textit{comparative} \textit{glossary} \textit{of} \textit{Cyp
year = {2004}
}
\end{styleBibliography}

\begin{styleBibliography}
———. 2008. „Lexical Aramaisms in Nomadic Arabic Vernaculars". In \textit{Between} \
112. Münster-Wien: LIT-Verlag.
\end{styleBibliography}

\begin{styleBibliography}
@@article{Brockelmann1928,
author = {Brockelmann, Carl},
journal = {\textit{Zeitschrift} \textit{für} \textit{Semitistik} \textit{und} \tex
pages = {109–34},
title = {„Deminutiv und Augmentativ im Semitischen"},
volume = {6},
year = {1928}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Brustad2000,
address = {Washington},
author = {Brustad, Kristen.},
```

```
publisher = {Georgetown University Press},
title = {\textit{The} \textit{Syntax} \textit{of} \textit{Spoken} \textit{{Arabi
year = {2000}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Cantineau1960,
address = {Paris},
author = {Cantineau, Jean.},
publisher = {Klincksieck},
title = {\textit{Études} \textit{de} \textit{linguistique} \textit{arabe}},
year = {1960}
}
\end{styleBibliography}

\begin{styleBibliography}
Coghill, Eleanor. 2015. „Parallels in the grammaticalization of Neo-Aramaic \tex
44. Wiesbaden: Harrassowitz.
\end{styleBibliography}

\begin{styleBibliography}
Contini, Ricardo. 1999. „Le substrat araméen en néo-arabe libanais : Préliminair
22. Padova: University Press.
\end{styleBibliography}

\begin{styleBibliography}
@@article{Corriente1969,
author = {Corriente, Federico},
journal = {\textit{Sefarad}},
pages = {3–11},
title = {„Qalqūl en Semitico"},
volume = {29},
year = {1969}
}
\end{styleBibliography}

\begin{styleBibliography}
```

```
@@book{Cowell1964,
address = {Washington D.C.},
author = {Cowell, Mark W.},
publisher = {Georgetown University Press},
title = {\textit{A} \textit{Reference} \textit{Grammar} \textit{of} \textit{Syrian}
year = {1964}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Diem1979,
author = {Diem, Werner},
journal = {\textit{Der} \textit{Islam}},
pages = {12–80},
title = {„{Studien} zur Frage des Substrats im Arabischen"},
volume = {56},
year = {1979}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Donner1981,
address = {Princeton},
author = {Donner, Fred McGraw.},
publisher = {Princeton University Press},
title = {\textit{The} \textit{early} \textit{{Islamic}} \textit{conquests}},
year = {1981}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Fassberg2010,
address = {Leiden \& Boston},
author = {Fassberg, Steven Ellis.},
publisher = {Brill},
title = {\textit{The} \textit{{Jewish}} \textit{Neo-Aramaic} \textit{dialect} \tex
year = {2010}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@@book{Féghali1918,
address = {Paris},
author = {Féghali, Michel.},
publisher = {Champion},
title = {\textit{Étude} \textit{sur} \textit{les} \textit{emprunts} \textit{syri
year = {1918}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Ferguson1969,
author = {Ferguson, Charles},
journal = {\textit{Word}},
pages = {114–19},
title = {„The /g/ in Syrian {Arabic}: {F}illing a gap in a phonological pattern"
volume = {25},
year = {1969}
}
\end{styleBibliography}

\begin{styleBibliography}
——. 1997. „Arabic baby talk". In \textit{Structuralist} \textit{studies} \texti
87. Leiden: Brill.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Fleisch1974,
address = {Beirut},
author = {Fleisch, Henri.},
publisher = {Dar el-Machreq},
title = {\textit{Études} \textit{d'arabe} \textit{dialectal}},
year = {1974a}
}
\end{styleBibliography}
```

\begin{styleBibliography}
———. 1974b. „Le changement ā > ō dans les sémitique de l'ouest et en arabe dialecta
50. Beyrouth: Dar el-Machreq.
\end{styleBibliography}


\begin{styleBibliography}
———. 1974c. „Le parler arabe de Kfar-Ṣghâb, Liban". In \textit{Études} \textit{d'ar
62. Beirut: Dar el-Machreq.
\end{styleBibliography}


\begin{styleBibliography}
@@book{Gardani2015,
address = {Berlin},
author = {Gardani, Francesco, Peter Arkadiev, und Nino Amiridze.},
publisher = {de Gruyter},
title = {\textit{Borrowed} \textit{morphology}},
year = {2015}
}
\end{styleBibliography}


\begin{styleBibliography}
Gazsi, Dénes. 2011. „Arabic-Persian Language Contact". In \textit{The} \textit{Sem
1021. Berlin, Boston: De Gruyter Mouton.
\end{styleBibliography}


\begin{styleBibliography}
@@book{Gralla2006,
address = {Wiesbaden},
author = {Gralla, Sabine.},
publisher = {Harrassowitz},
title = {\textit{Der} \textit{arabische} \textit{{Dialekt}} \textit{von} \textit{Na
year = {2006}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Grigore2007,
address = {Bucharest},

```
author = {Grigore, George.},
publisher = {Editura Universităţii din Bucureşti},
title = {\textit{L'arabe} \textit{parlé} \textit{à} \textit{Mardin:} \textit{Mon
year = {2007}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Halasi-Kun1969,
author = {Halasi-Kun, Tibor},
journal = {\textit{Archivum} \textit{Ottomanicum}},
pages = {14–91},
title = {„The Ottoman elements in the Syrian dialects"},
volume = {1},
year = {1969}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{———1973,
author = {———},
journal = {\textit{Archivum} \textit{Ottomanicum}},
pages = {17–95},
title = {„The Ottoman Elements in the Syrian Dialects, II"},
volume = {5},
year = {1973}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{———1982,
author = {———},
journal = {\textit{Archivum} \textit{Ottomanicum}},
pages = {117–267},
title = {„The Ottoman elements in the Syrian dialects, II (continuation)"},
volume = {7},
year = {1982}
}
```

\end{styleBibliography}

\begin{styleBibliography}
Holes, Clive. 2002. „Non-Arabic Semitic elements in the Arabic dialects of eastern
79. Wiesbaden: Harrassowitz.
\end{styleBibliography}

\begin{styleBibliography}
——. 2007. „Colloquial Iraqi Arabic". In \textit{Languages} \textit{of} \textit{Ira}
34. Bagdad: British School of Archaeology in Iraq.
\end{styleBibliography}

\begin{styleBibliography}
@@book{——.2016,
address = {Leiden; Boston},
author = {——.},
publisher = {Brill},
title = {\textit{Dialect,} \textit{Culture,} \textit{and} \textit{Society} \texti
year = {2016}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Hopkins1995,
author = {Hopkins, Simon},
pages = {37–49},
title = {„\textit{ṣarār} ‚pebbles' – A Canaanite Substrate Word in Palestinian {Ar
volume = {30},
year = {1995}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{——1997,
author = {——},
journal = {\textit{Journal} \textit{of} \textit{Semitic} \textit{Studies}},
pages = {23–32},
title = {„On the construction šmēh l-gabrā 'the name of the man' in Aramaic"},

```
volume = {42},
year = {1997}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Jastrow1979,
author = {Jastrow, Otto},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguist
pages = {36–76},
title = {„Zur arabischen {Mundart} von Mossul"},
volume = {2},
year = {1979}
}
\end{styleBibliography}

\begin{styleBibliography}
———. 1990. „Die arabischen Dialekte der irakischen Juden". In \textit{XXIV.} \te
205. Stuttgart: Steiner.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Kahane1958,
address = {Urbana},
author = {Kahane, Henry Romanos, Renée Kahane, und Andreas Tietze.},
publisher = {Univ. of Illinois Press},
title = {\textit{The} \textit{Lingua} \textit{Franca} \textit{in} \textit{the} \
year = {1958}
}
\end{styleBibliography}

\begin{styleBibliography}
Khoury, Philipp S. 1987. \textit{Syria} \textit{and} \textit{the} \textit{French
\end{styleBibliography}

\begin{styleBibliography}
Krebernik, Manfred. 2008. „Von Gindibu bis Muḥammad: Stand, Probleme und Aufgabe
79. Wiesbaden: Harrassowitz.
```

```
\end{styleBibliography}

\begin{styleBibliography}
@@book{Lentin2018,
address = {Oxford},
author = {Lentin, Jérôme.},
publisher = {Oxford University Press},
title = {„The Levant". {I}n \textit{The} \textit{Historical} \textit{Dialectology}
year = {2018}
}
\end{styleBibliography}

\begin{styleBibliography}
Lucas, Christopher. 2015. „Contact-induced language change". In \textit{The} \text
36. London: Routledge.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Magidow2013,
address = {Austin},
author = {Magidow, Alexander.},
publisher = {University of Texas},
title = {„Towards a Sociohistorical Reconstruction of Pre-{Islamic} Arabic Dialect
year = {2013}
}
\end{styleBibliography}

\begin{styleBibliography}
Masliyah, Sadok. 1996. „Four Turkish suffixes in Iraqi Arabic: -li, -lik, -siz and
300.
\end{styleBibliography}

\begin{styleBibliography}
@@article{———1997,
author = {———},
pages = {68–88},
title = {„The Diminutive in Spoken Iraqi {Arabic}". \textit{Zeitschrift} \textit{f
volume = {33},
```

```
year = {1997}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Muraoka2005,
address = {Wiesbaden},
author = {Muraoka, Takamitsu.},
publisher = {Harrassowitz},
title = {\textit{Classical} \textit{Syriac}. 2., revised edition. {P}orta lingua
year = {2005}
}
\end{styleBibliography}


\begin{styleBibliography}
Neishtadt, Mila. 2015. „The lexical component in the Aramaic substrate of Palest
310. Studies in Semitic Languages and Linguistics 82. Leiden: Brill.
\end{styleBibliography}


\begin{styleBibliography}
@@book{Nöldeke1904,
address = {Straßburg},
author = {Nöldeke, Theodor.},
publisher = {Trübner},
title = {\textit{{Neue}} \textit{Beiträge} \textit{zur} \textit{semitischen} \te
year = {1904}
}
\end{styleBibliography}


\begin{styleBibliography}
@@book{Owens2006,
address = {Oxford},
author = {Owens, Jonathan.},
publisher = {Oxford University Press},
title = {\textit{A} \textit{linguistic} \textit{history} \textit{of} \textit{{Ar
year = {2006}
}
\end{styleBibliography}
```

\begin{styleBibliography}
Palva, Heikki. 2009. „From qəltu to gilit: Diachronic notes on linguistic adaptatic
40. Leiden \& Boston: Brill.
\end{styleBibliography}

\begin{styleBibliography}
Pat-El, Na'ama. 2017. „Digging up archaic features: "Neo-Arabic" and comparative S
75. Leiden: Brill.
\end{styleBibliography}

\begin{styleBibliography}
Perry, John. 2007. „Persian morphology". In \textit{Morphologies} \textit{of} \text
1019. Winona Lake, Ind: Eisenbrauns.
\end{styleBibliography}

\begin{styleBibliography}
Procházka, Stephan. 2002a. „Contact phenomena, code-copying, and code-switching in
39. Rabat: Amapatril.
\end{styleBibliography}

\begin{styleBibliography}
@@book{———.2002,
address = {Wiesbaden},
author = {———.},
publisher = {Harrassowitz},
title = {\textit{Die} \textit{arabischen} \textit{{Dialekte}} \textit{der} \textit
year = {2002b}
}
\end{styleBibliography}

\begin{styleBibliography}
———. 2004. „Unmarked feminine nouns in modern Arabic dialects". In \textit{Approach
82. Leiden: Brill.
\end{styleBibliography}

\begin{styleBibliography}
———. 2013. „Traditional boatbuilding: Two texts in the Arabic dialect of the island

```
88. Wiesbaden: Harrasowitz.
\end{styleBibliography}

\begin{styleBibliography}
@@book{———.2018,
address = {Oxford},
author = {———.},
publisher = {Oxford University Press},
title = {„The {Northern} Fertile Crescent". {I}n \textit{The} \textit{Historical
year = {2018}
}
\end{styleBibliography}

\begin{styleBibliography}
Procházka, Stephan, und Ismail Batan. 2016. „The Functions of Active Participles
} \textit{Bucharest,} \textit{2015}, herausgegeben von George Grigore und Gabriel
66. Bucharest: Editura Universității din București.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Procházka-Eisl2018,
address = {„A suffix on the move},
author = {Procházka-Eisl, Gisela.},
publisher = {The history and development of the Turkish suffix -\textit{ci} in A
year = {2018}
}
\end{styleBibliography}

\begin{styleBibliography}
Prunet, Jean-François, und Ali Idrissi. 2014. „Overlapping morphologies in Arabi
92. Amsterdam/Philadelphia: John Benjamins.
\end{styleBibliography}

\begin{styleBibliography}
@@article{Reinkowski1995,
author = {Reinkowski, Maurus},
journal = {\textit{Folia} \textit{Orientalia}},
pages = {89—115},
```

```
title = {„Eine phonologische Analyse türkischen Wortgutes im Bagdadisch-Arabischen
volume = {31},
year = {1995}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Retsö2006,
address = {Leiden},
author = {Retsö, Jan.},
publisher = {Brill},
title = {„Aramaic/Syriac Loanwords". {H}erausgegeben von Kees Versteegh. \textit{En
year = {2006}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Rubin2005,
address = {Winona Lake},
author = {Rubin, Aaron D.},
publisher = {Eisenbrauns},
title = {\textit{Studies} \textit{in} \textit{{Semitic}} \textit{grammaticalizatior
year = {2005}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Sabuni1980,
address = {Frankfurt am Main},
author = {Sabuni, Abdulghafur.},
publisher = {Peter Lang},
title = {\textit{{Laut}-} \textit{und} \textit{Formenlehre} \textit{des} \textit{a
year = {1980}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Seifart2013,
```

```
address = {\textit{Leipzig},
author = {Seifart, Frank.},
note = {//afbo.info.},
publisher = {} \textit{Max} \textit{Planck} \textit{Institute} \textit{for} \tex
title = {„AfBo: {A} world-wide survey of affix borrowing"},
year = {2013}
}
\end{styleBibliography}

\begin{styleBibliography}
Shields, Sarah. 2004. „Mosul questions: economy, identity, and annexation". In \
60. New York: Columbia University Press.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Singer1984,
address = {Berlin/New York},
author = {Singer, Hans-Rudolf.},
publisher = {de Gruyter},
title = {\textit{{Grammatik}} \textit{der} \textit{arabischen} \textit{Mundart}
year = {1984}
}
\end{styleBibliography}

\begin{styleBibliography}
@@article{Souag2017,
author = {Souag, Lameen},
journal = {\textit{Zeitschrift} \textit{für} \textit{arabische} \textit{Linguist
pages = {45–70},
title = {„Clitic doubling and language contact in {Arabic}"},
volume = {66},
year = {2017}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Talay1999,
address = {Wiesbaden},
```

```
author = {Talay, Shabo.},
publisher = {Harrassowitz},
title = {\textit{Der} \textit{arabische} \textit{{Dialekt}} \textit{der} \textit{K}
year = {1999}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Trudgill1986,
address = {Oxford},
author = {Trudgill, Peter.},
publisher = {Blackwell},
title = {\textit{Dialects} \textit{in} \textit{contact}},
year = {1986}
}
\end{styleBibliography}

\begin{styleBibliography}
Weninger, Stefan. 2011. „Aramaic-Arabic Language Contact". In \textit{The} \textit{
55. Berlin: De Gruyter Mouton.
\end{styleBibliography}

\begin{styleBibliography}
@@book{Wilkins2010,
address = {Leiden},
author = {Wilkins, Charles L.},
publisher = {Brill},
title = {\textit{Forging} \textit{urban} \textit{solidarities:} \textit{Ottoman} \
year = {2010}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Wilmsen2014,
address = {Oxford},
author = {Wilmsen, David.},
publisher = {Oxford University Press},
title = {\textit{{Arabic}} \textit{indefinites,} \textit{interrogatives,} \textit{
```

```
year = {2014}
}
\end{styleBibliography}

\begin{styleBibliography}
@@book{Woidich2006,
address = {Wiesbaden},
author = {Woidich, Manfred.},
publisher = {Harrassowitz},
title = {\textit{Das} \textit{Kairenisch-Arabische:} \textit{eine} \textit{{Gram
year = {2006}
}
\end{styleBibliography}
```

## Abbreviations

## Acknowledgements

# Chapter 2

# Contact-induced change in Khuzestan Arabic

Betina Leitner

University of Vienna

Khuzestan Arabic is an Arabic variety spoken in the southwestern Iranian province of Khuzestan. It has been in contact with (Modern) Persian since the arrival of Arab tribes in the region before the rise of Islam. Persian is the socio-politically dominant language in the modern state of Iran and has influenced the grammar of Khuzestan Arabic on different levels. The present article discusses phenomena of contact-induced change in Khuzestan Arabic and considers their limiting factors.

## 1 Current state and historical development

### 1.1 Historical development

Arab settlement in Iran preceded the Arab destruction of the Sasanian empire with the rise of Islam. Various tribes, such as the Banū Tamīm, had settled in Khuzestan prior to the arrival of the Arab Muslim armies (**Daniel1986**). In the centuries after the spread of Islam in the region, large groups of nomads from the Ḥanīfa, Tamīm, ʕAbd-al-Qays, and other tribes crossed the Persian Gulf and occupied some of the territories of southwestern Iran (**Oberling1986**). The Kaʕb, still an important tribe in the area,[1] settled there at the end of the 16th century (**Oberling1986**). During the succeeding centuries many more tribes moved from southern Iraq into Khuzestan. This has led to a considerable increase of Arabic speakers in the region, which until 1925 was called Arabistan (see **Gazsi2011**: 1020; and Gazsi, this volume). Today Khuzestan is one of the 31 provinces of the Islamic Republic of Iran, situated in the southwest at the border to Iraq.

---

[1]Cf. **Oberling1986** for an overview of the Arab tribes in Khuzestan.

There has been considerable movement to and from Iraq, to Kuwait, Bahrain, and Syria, and from villages into towns. Many of these migrations were a consequence of the Iran–Iraq war (1980–88), but some were due to socio-economic reasons. The settlement of Persians in the region over the past decades (**Gazsi2011**) is another important factor in its demographic history. From the early 20th century on Khuzestan has attracted international, especially British, interest because of its oil resources.

## 1.2 Current situation of Arabs in Khuzestan

Information about the exact number of Arabic-speaking people in Iran, and especially in Khuzestan, is hard to find. Estimates in the 1960s of the Arabic-speaking population in Iran ranged from 200,000 to 650,000 (**Oberling1986**). Today it is estimated that around 2 to 3 million Arabs live in Khuzestan (**MatrasShabibi2007**: 137; **Gazsi2011**: 1020).

Many Arabs and Persians living in Khuzestan work in the sugar cane or oil industries, but few hold white-collar or managerial positions (de **Planhol1986**: 55–56). This is one of the reasons why many Arabs in Khuzestan feel strongly disadvantaged in society and politics in comparison to their Persian neighbors.[2]

## 2 Language contact in Khuzestan

Currently, the main and most influential language in contact with Khuzestan Arabic is the Western Iranian language Persian. Among the other (partly historically) influential languages in the region the most prominent are English, Turkish/Ottoman (cf. **Ingham2005**), and Aramaic (see Procházka, this volume).

Persian and different forms of Arabic share a long history of contact in the region of Khuzestan, implying a long exchange of language material in both directions.

Khuzestan Arabic belongs to the Bedouin-type South-Mesopotamian *gələt*-dialects.[3] Therefore, it shows great similarity to such Iraqi dialects as Baṣra Arabic and Muslim Baghdadi Arabic as well as to other Bedouin dialects and the Gulf

---

[2]The Khuzestan Arabic terms for the Persian people and their language are *ʕaǧam* 'non-Arab; Persian (people and language)', and *əl-ǧamāʕa* 'Persians'. Both are often used pejoratively.

[3]There is as yet no comprehensive grammar of the dialects of Khuzestan. The main source of information on these dialects is the collection of data made in the 1960s by the arabist and linguist Bruce Ingham (2007; 1976; 1973). The article by Yaron Matras and Maryam Shabibi, "Grammatical borrowing in Khuzistani Arabic" (??), is based on Shabibi's unpublished dissertation "Contact-induced grammatical changes in Khuzestani Arabic" (??).

dialects like Bedouin Bahraini Arabic – that is, the Arabic spoken by the Sunni Arab population descended from Najd.

The dialects of Khuzestan can be considered "peripheral" dialects of Arabic because they are spoken in a country where Arabic is not the language of the majority population and is not used in education and administration. Therefore, there is practically no influence of Modern Standard Arabic. However, because it shares a long geographically-open border with Iraq, Khuzestan is not isolated from the Arabic-speaking world. Moreover, since around 2000 it has had access to Arabic news, soaps, etc. via satellite TV. Intra-Arabic contact is limited to the linguistically very similar (southern) Iraqi dialects[4] through, for example, religious visits to Kerbala.

Persian is the only official language in Iran and the only language used in education and is sociolinguistically and culturally dominant, especially in the domains of business and administration. Persian consequently enjoys high prestige in society. For Persian speakers, and sometimes also for Khuzestan Arabic speakers, the Khuzestan Arabic varieties have very low prestige and are not associated with the highly prestigious Arabic of the Quran, which is taught at schools. Khuzestan Arabic speakers who acquire Khuzestan Arabic as a first language usually acquire Persian at school. Later, the opportunities for Khuzestan Arabic speakers to use Persian are restricted to certain social settings outside the family, e.g. school, work (employment in a large company would probably require communication in Persian), contact with Persian friends, or through the Persian media.

Accordingly, the command of Persian or the degree of bilingualism among Khuzestan Arabic speakers varies greatly due to such factors as level of education, affiliation, age, gender, and urban or rural environment. The older generation and women have far less access to education and jobs and consequently less contact with people outside the family, which implies less exposure to contact situations and a lower degree of bilingualism. Among the younger generation we notice a certain intentional re-inforcement of Arabic words alongside a resistance to recognizable Persian lexical borrowings plus a preference for the Arabic over the Persian names for the cities in Khuzestan. This of course goes in line with nationalist ideas and the separatist movement taking place in present-day Khuzestan and also shows the impact of intentionality in language contact situations.

---

[4]Khuzestan Arabic is often differentiated from its neighboring Iraqi dialects by the number of Persian borrowings that are employed (**Gazsi2011**). Although the greatest influence has occurred in lexicon, Persian influence also extends to grammar (see below).

In sum, one might find very different degrees of Persian influence among the speakers of Khuzestan Arabic (cf. **MatrasShabibi2007**: 147). For that reason, all statements on Persian–Khuzestan Arabic contact phenomena must be seen in relation to the above factors, which are decisive for any speaker's command of Persian.

# 3  Contact-induced changes in Khuzestan Arabic

## 3.1  General remarks

[5] The main aim of the present chapter is to highlight the most striking phenomena and trends in Khuzestan Arabic language change due to contact with Persian.

All phenomena of contact-induced change in Khuzestan Arabic can be considered as transfer of patterns or matter[6] from the source language (SL) Persian to the recipient language (RL) Khuzestan Arabic under RL agentivity (i.e. borrowing rather than imposition). The agents of transfer are cognitively dominant in the RL Khuzestan Arabic, the agents' L1. Even though Persian is generally acquired during childhood and today is spoken by most speakers, it usually is the speakers' L2. Cases of convergence (cf. Lucas, 2015: 530–531) are possible in the present contact situation among speakers with a very high (L1-like) command of Persian, for example university students. But of course it is hard to draw an exact line between L1 and L2 proficiency and thus between convergence and borrowing (cf. Lucas, 2015: 531).

## 3.2  Phonology

As in other Bedouin Arabic dialects, the presence of the phonemes /č/ and /g/ is ultimately the result of internal development from original *k and *q, rather than borrowing from Persian (see Procházka, this volume).

The phoneme /p/, e.g. *perde* 'curtain' < P *parde*, is also common in all Iraqi dialects and probably emerged in this region due to contact with Persian and Kurdish (see Procházka, this volume).

---

[5]The data used for the present analysis was collected mainly in Aḥwāz, Muḥammara (Khorramshahr), Ḥamīdīye and Ḫafağīye (Susangerd) in 2016. The male and female informants were bilingual as well as monolingual Khuzestan Arabic speakers from 25 to over 70 years old.

[6]According to **Sakel2007**, matter replication is the replication of "morphological material and its phonological shape".

An interesting phonological feature of Khuzestan Arabic is that /ɣ/ often reflects etymological *q[7], which is otherwise realized as /g/ and /ǧ/. /ɣ/ < *q mainly occurs in Persian borrowings ultimately of Arabic origin, e.g. *ɣisma* 'part, section', *taṣdīɣ* 'driving licence', *yifruɣ* 'it differs', *taɣrīban* 'approximately', and *ɣidra* 'power'. This feature is most probably a transfer from Persian,[8] in which both *q and *ɣ in Arabic loanwords are always pronounced /ɣ/[9] (**MatrasShabibi2007**: 138).

Lexical borrowings are often adapted to Arabic phonology. For example, Khuzestan Arabic speakers of the older generation usually pronounce the phoneme /p/ as /b/; and /s/ in Persian loanwords is sometimes pronounced /š/, e.g. *berde* 'curtain' < P *parde*, and *šebzi* 'herb stew' < P *sabzi*.

Negative structures bear stress on the first syllable,[10] e.g. KhA *mā arūḥ* 'I don't go'. This is a feature shared with some Persian and Turkish varieties and other North East Arabian dialects (**Ingham2005**: 178–179). This common phonological characteristic therefore seems to be a Sprachbund phenomenon of the Mesopotamian region, which reflects the long history of contact and migration across language boundaries due to trade, war, shared cultural practices, nomadism, etc. (**Winford2003**: 70–74). Though the directions and mechanisms of borrowing within the languages of a Sprachbund are often hard to categorize (**Winford2003**), we can probably assume that Khuzestan Arabic, being spoken by a minority group, has borrowed and adapted this phonological stress pattern under RL agentivity.

## 3.3 Morphosyntax

### 3.3.1 Replication of Persian phrasal verbs

The replication of phrasal verbs is a contact phenomenon also found in the Arabic varieties of Turkey (**Grigore2007**: 157–159; and Procházka, this volume). As shown in examples (1), (2), (3) and (4), Khuzestan Arabic replicates Persian phrasal verbs by substituting the Persian light verbs with Khuzestan Arabic equivalents and directly replicating the Persian nouns (cf. **MatrasShabibi2007**: 142). As examples (3) and (4) show, the Persian nouns are sometimes adapted morpho-phonologically.

---

[7]This phenomenon is also documented for the Arabic dialects of Kuwait, Qatar, and the United Arabic Emirates (**Holes2016**: 54, fn. 5).

[8]Contrast Holes (2016: 53–54), who explains the /ɣ/–/q/ merger among the Najd-descendent Bahraini Arabic speakers as internal development.

[9]In Modern Persian the phoneme /ɣ/ has two allophones, [ɢ] and [ɣ] (**Majidi1986**: 58–60).

[10]**Ingham1991** describes this phenomenon also for Khuzestan Arabic wh-interrogatives and prepositions.

The noun in example (1) is Arabic in its origins but its usage in a phrasal verb construction with a new meaning is a Persian innovation.

(1)  a.  KhA (own data)
        ṭəgg        muḫḫ
        hit.PRF.3SG.M brain

    b.  Persian
        muḫḫ zadan
        brain  hit.INF
        'to brainwash/convince someone'

(2)  a.  KhA (own data)
        kaḏ̣ḏ̣        irād
        take.PRF.3SG.M nagging

    b.  Persian
        īrād     gereftan
        nagging take.INF
        'to pick at someone'

(3)  a.  KhA (own data)
        sawwa          ʔōmāde
        make.PRF.3SG.M ready

    b.  Persian
        āmāde kardan
        ready  make.INF
        'to prepare sth.'

(4)  a.  KhA (own data)
        ṭalaʕ  ɣabūli[11]
        come out.PRF.3SG.M acceptance

    b.  Persian
        ɣabūl     šodan
        acceptance become.INF
        'to pass (an exam), be accepted'

---

[11]The final -i in ɣabūli probably originates from the Persian indefiniteness marker -i (see **Majidi1990**: 309–314) and has become part of this word in Khuzestan Arabic, so that ɣabūli is monomorphemic.

The pattern for phrasal verbs – transferred into the RL Khuzestan Arabic under RL agentivity – provides Khuzestan Arabic with an easy way to convert foreign nouns into verbs.

The pattern is adapted according to Arabic syntactic rules: (i) the verb is moved into the initial position; and (ii) a direct object is introduced between verb and nominal element (post-verbally) (see examples (5)–(6)). In Persian, however, the verb always remains in final position following the nominal element and a direct object would be introduced before both elements (see e.g. **Majidi1990**: 447–448).

(5)  a.  KhA (own data)
    ṭəggi       ɣandart-i    wāks!
    hit.IMP.2SG.F shoe-OBL.1SG wax

  b.  Persian
    kafš-am-o       wāks be-zan!
    shoe-OBL.1SG-ACC wax  IMP-hit.PRS

    'Polish my shoes!'

(6)  a.  KhA (own data)
    yṭəggūn      əṭ-ṭamāṭe  rande.
    hit.IMPF.3PL.M DEF-tomato grater

  b.  Persian
    gūǧe        farang-i rō   rande         mī-zanan.
    tomato-INDEF ACC      grater PROG-hit.PRS.3PL.M

    'They grate some tomato.'

This structure has become productive in Khuzestan Arabic. For example, in the phrasal verb *ṭəgg dabbe* 'to cheat' (lit. 'to hit a water canister') both the verb and noun are taken from Khuzestan Arabic and only the construction's pattern remains Persian.

## 3.4  Syntax

### 3.4.1  Definiteness marking

Matras & Shabibi (2007: 141–142) see Khuzestan Arabic relative clauses without definite heads as evidence for the decline of overt definiteness marking in Khuzestan Arabic based on the Persian model with generally unmarked definiteness, e.g. *mara lli šiftū-ha ḫābarat* 'The woman that you saw called.' (2007: 142). However, this pattern is also documented in Arabic dialects which have had no contact with Persian (Pat-El, 2017: 454–455; cf. Procházka, 2018: 269).

Matras & **Shabibi2007** further postulate that the Persian Ezāfe pattern[12] in adjectival attribution is replicated in Khuzestan Arabic. According to their theory, the construct state marker −*t* (with an indefinite head) and/or the definite article (of the attribute) are reanalysed as markers of attribution matching the Persian Ezāfe marker -(*y*)*e*, e.g.

(7)  a.  KhA (**MatrasShabibi2007**: 140)
          ğazīra-t    l-ḫaḍra
          *island*-CON DEF-*green.*F
     b.  Persian
          ğazīre-ye sabz
          *island*-EZ *green*

     'the green island'

However, this pattern is also observed in other modern Arabic dialects which have not been exposed to Persian influence as well as in older forms of Arabic.[13]

Consequently, it is highly unlikely that this phenomenon has developed due to Persian influence, although it cannot be ruled out for that contact with Persian has fostered the preservation of this apparently old feature.

### 3.4.2  Word order changes

Khuzestan Arabic shows no changes due to contact in basic word order.[14] The only attested word order changes concern the position of the verbs *čān* 'to be' and *ṣār* 'to become', both of which can appear in final position as an unmarked construction. This sentence-final position in no case functions as the default and is in fact less frequent than its non-final position.[15] *čān* or *ṣār* in final position are never stressed.

---

[12]See e.g. Ahadi (2001: 103–109) for the usage of the Persian Ezāfe.

[13]See Pat-El (2017: 445–449) for numerous examples from different modern Arabic dialects, Middle Arabic, Pre-Islamic Arabic, Quran Arabic, and other Central Semitic languages; see also Retsö (2009: especially 21–22); Procházka (2018: 267–269) also proves that this is an old feature already found in Old Arabic and points out that it is mainly found among dialects which are spoken in regions with no or only marginal influence from Modern Standard Arabic.

[14]**Ingham1991** states that in Khuzestan Arabic neither VSO nor SVO word order is particularly dominant. Matras and **Shabibi2007** postulate that the usage of OV order in Khuzestan Arabic is increasing as "the beginning of a shift in word order" on the basis of the Persian type, where OV prevails. In both of their examples the objects are topicalized (with pronominal resumption), which is a common phenomenon in spoken Arabic (**Brustad2000**: 330–333; 349) and as such not the result of Persian influence.

[15]In my data, *čān* appears 17 of 137 times in sentence-final position, *ṣār* 11 of 165 times. The additional examples are taken from my questionnaire.

2 Contact-induced change in Khuzestan Arabic

The sentence-final position of *čān* or *ṣār* is likely a pattern replication of the Persian model, i.e. sentences with final *būdan* 'to be' or *šodan* 'to become'.

(8)  a.  KhA, ʕAbbādān (own data)
šuɣul-hum        b-əl-bandar čān.
work-OBL.3PL.M in-DEF-port be.PRF.3SG.M

b.  Persian
kār-ešūn        tū-ye bandar būd.[16]
job-OBL.3PL in-EZ port        be.PST.3SG

'Their job was at the port.'

(9)  a.  KhA, Muḥammara (own data)
əğdād-i                mallāk-īn čānaw.
grandparents-OBL.1SG owner-PL be.PRF.3PL.M

b.  Persian
pedarbozorgā-m        mālek būdan.
grandparents-OBL.1SG owner be.PST.3PL

'My grandparents were owners [of land].'

(10)  a.  KhA, Aḥwāz (own data)
hassa šway l-māy        bārəd ṣār.
now   a_bit DEF-water cold   become.PRF.3SG.M

b.  Persian
alʔān yekam ʔāb        sard šod.
now   a_bit    water cold become.PST.3SG

'The water has become a bit cold now.'

The next example might show a tendency to use a copula with human subjects, expressed with the verb *ṣār* 'to become':

(11)  a.  KhA, Aḥwāz (own data)
əhye mart uḵ-ūy                əṣṣīr.
3SG.F wife  brother-OBL.1SG COP.IMPF.3SG.F

b.  Persian
ūn  zan  dādāš-am-e.
3SG wife brother-OBL.1SG-COP.PRS.3SG

'She is the wife of my brother.'

---

[16]All Persian translations are given in the modern spoken (Tehrani) variety of Persian.

(12) In the Khuzestan Arabic construction for pluperfect tense, *čān* can also appear in sentence-final position, after the active participle. This construction is very likely a direct transfer of the Persian structure, in which the auxiliary *būdan* also follows the participle.[17]

    a. KhA, Aḥwāz (own data)
       lamman iyēna     l-əl-bīət,     əhma mākl-īn     čānaw.
       when    come.PRF.1PL to-DEF-house 3PL.M eat.PTCP-M.PL be.PRF.3PL.M

    b. Persian
       vaɣti-ke   mā bargaštīm      ḫūne, ūnhā ɣazā-ro   ḫorde
       when-REL 1PL come_back.PST.1PL home 3PL    food-ACC eat.PTCP
       būdan.
       be.PST.3PL

       'When we came home, they had (already) eaten.'

This word order change has probably been triggered by the high frequency in speech of Persian sentences with forms of *būdan* in final position. **Lucas2012** explains the usage of foreign patterns as the result of the human cognitive tendency to minimize the high processing efforts associated with the extensive use of two languages.[18]

*čān* is also used in sentence-final positions after the main verb in the imperfective in Khuzestan Arabic constructions expressing the continuous past. In Persian, the continuous past is formed without a sentence-final *būdan*.[19] This case is not a direct transfer of the Persian pattern, but perhaps a construction analogous to the pluperfect and other Persian forms with *būdan* in final position.

(13)   a. KhA, Aḥwāz (own data)
       hāde     ham mən zuɣur    yəštəɣəl     čān.
       DEM.SG.M also from childhood work.IMPF.3SG.M be.PRF.3SG.M

---

[17]Matras and Shabibi (2007: 142–143) describe the use of this construction as a change in the Khuzestan Arabic tense system. However, the pattern *kān* + active participle is also commonly used in other Arabic dialects to express pluperfect meaning or to describe completed actions which have an impact on the present, see for example Denz (1971: 92–94; 115–116) for Iraqi (Kwayriš) and **Grotzfeld1965** for Syrian Arabic.

[18]Connections between units of a neural network associated with certain syntactic patterns can be strengthened from repeated exposure to and use of that pattern (**Lucas2012**). Hence, the employment of a Persian syntactic structure in Khuzestan Arabic needs less processing effort because the same strengthened neural network is activated.

[19]The Persian continuous past is formed with the particle *mī* prefixed to the simple past of the respective main verb and can (for the progressive form) be preceded by the simple past of *dāštan* 'to have': e.g. (*dāšt*) *mī-raft* 'he was going' (**Majidi1990**: 232, 235).

b. Persian

īn      ham az   kūdeki     kār  mī-kard.
DEM.SG also from childhood work PROG-do.PST.3SG

'This one has also been working from childhood on.'

(14)   The following example shows both syntactic variants in one sentence:

a. KhA, Muḥammara (own data)

umm-i        čānat        təthaǧǧab,       eh, əb-zamān əš-šāh,
mother-OBL.1SG be.PRF.3SG.F veil.IMPF.3SG.F yes in-time    DEF-shah

bass titbawwaš     čānat.
only veil.IMPF.3SG.F be.PRF.3SG.F

b. Persian

mādar-am (dāšt)      naqāb mī-zad,        āre, dar zamān-e
mother-1SG (have.PST.3SG) veil PROG-hit.PST.3SG yes in time-EZ

šāh,  hamīše naqāb mī-zad.
shah always veil   PROG-hit.PST.3SG

'My mother was veiling her face (with a *būšīye*)[20], yes, during the
times of the shah, she was always veiling her face.'

Because all the above examples equally work with *čān/ṣār* in non-final posi-
tion, the process of pattern replication or word order change in Khuzestan Arabic
is still ongoing. Indeed, all informants, when asked for the correct structure in
the above examples, preferred the verb *čān* in non-final position.[21]

Lucas (2015: 530–531) explains the word order changes (from VSO to SOV) in
Bukhara Arabic (cf. **Ratcliffe2005**: 143–144; and **Versteegh2010**: 639) as a result of
convergence[22] with Uzbeki. Although a clear division between convergence and
borrowing is hard to make, I consider the contact-induced word order changes in
Khuzestan Arabic to be instances of borrowing because most speakers are clearly
native speakers of, and therefore dominant in, Khuzestan Arabic only.

### 3.4.3  *ḫōš* preceding verbs and nouns

In Persian, *ḫōš* 'good, well' is used as a prefixed (lexicalised) element preceding
some nouns and verbs to coin compound adjectives, nouns, and verbs (**Majidi1990**:

---

[20] *būšīye* or *pūšīye* 'veil' is also documented for Iraqi Arabic (**WoodheadBeene1967**: 53).

[21] My informants from Baghdad considered all constructions with *čān* in final position to be
wrong. However, this structure is used in Baṣra Arabic (Qasim Hassan, pers. comm., **January2018**).

[22] **Lucas2015** defines convergence as being changes made to a language under the agentivity of
speakers who are native speakers of both the SL and the RL.

411, 413): e.g. *ḫōš-andām* 'handsome' (< P *andām* 'shape; body'), *ḫōš-nevīs* 'calligrapher' (< P present stem *nevīs-* 'to write'), and *ḫōš-āmadan* 'to please' (< P *āmadan* 'to come').

Khuzestan Arabic has borrowed some of these Persian compound adjectives: e.g. *ḫōš-bū* 'nice-smelling' (< P *bū* 'smell, scent'), *ḫōš-tīp* 'handsome' (< P *tīp* 'type'), and *ḫōš-aḫlāq* '(with) good manners' (< P *aḫlāq* 'decency; ethics, morality', pl. of *ḫolq* 'character, nature'). However, in Khuzestan Arabic *ḫōš* has become productive and occurs as an adjective used attributively preceding nouns, but not agreeing in gender or number with them, e.g. *ḫōš walad* 'a good boy', *ḫōš əbnayya* 'a good girl', *ḫōš banāt* 'good girls', *ḫōš əwlād* 'good kids'.[23]

## 3.5 Lexicon

### 3.5.1 Lexical transfer

The greatest influence from Persian on Khuzestan Arabic has occurred in lexicon. Many Persian lexemes were borrowed generations ago. The most frequently borrowed elements are nouns denoting cultural or technological innovations which have filled lexical gaps in the RL Khuzestan Arabic. Verbs, adverbs, adjectives, and many discourse particles have also been borrowed from the SL Persian.

The majority of the examples below are cases of transfer of morpho-phonological material (matter) and semantic meaning (pattern) under RL agentivity.

Many of the Persian borrowings have been phonologically and morphologically integrated into the RL. For instance, for many borrowed Persian nouns Arabic internal plural forms are created, e.g. *ḫətākīr* 'ball-point pens' (sg. *ḫətkār* < P *ḫod-kār* 'ball-point pen'), or *banādər* 'ports' (sg. *bandar* < P *bandar* 'port').

Again, the borrowing of foreign (L2) elements into the speakers' L1 might be explained by the human cognitive tendency to minimize the processing effort in lexical selection between two languages (**Lucas2012**: 291; see §??). So if a certain Persian word is frequently used and often heard (for example at school), the connections of a neural network associated with this word are strengthened (**Lucas2012**), which makes it easier to employ the word in one's L1.

### 3.5.2 Semantic fields

The following illustrative list of Persian loans in Khuzestan Arabic shows the most important semantic fields of lexical borrowing.

---

[23]This construction is also found in Iraqi Arabic (cf. **Erwin1963**: 256), which might prove that the element *ḫōš* is an older borrowing.

- Administration/Military: *farmāndāri* 'governorship' < P *farmāndāri*; *čārra* 'crossroad' < P *čahār-rāh*; *sarbāz/šarbāz* 'soldier' < P *sarbāz*.

- Agriculture: *ʕalafkoš* 'pesticide (lit. weed-killer)' < P *ʕalaf-koš*; *kūd* 'dung' < P *kūd*.

- Dress/Textiles: *dāmen* 'skirt' < P *dāman*; *šīəla* 'head covering' < P *šāl* 'Kashmir shawl' (**Ingham2005**).

- Education: *klāṣ* 'class, grade' < P *kelās*; *ḫətkār* 'biro' < P *ḫod-kār*; *dānišga* 'university' < P *dānišgāh*.

- Food: *ǧaʕfari* 'parsley' < P *ǧaʕfari*; *češmeš* 'raisins' < P *kešmeš*; *šalɣam* 'turnip' < P *šalɣam*; *serke* 'vinegar' < P *serke*.

- Material culture: *šīše* 'bottle' < P *šīše*; *ǧām* '(window) glass' < P *ǧām*; *tīye* 'blade' < P *tīye*; *yəḥčāle* 'refrigerator' < P *yahčāl*; *sīm buksel* 'towrope' < P *sīm-e boksol*; *perde/berde* 'curtain' < P *parde*; *gīre* 'hair barrette' < P *gīre-ye sar/mūy*; *mīz* 'table' < P *mīz*; *darīše* 'window' < P *darīče*; *pənǧara* 'window' < P *panǧare*.

- Other: *ɣīme* 'price' < P *ɣīmat*; *bandar* 'port' < P *bandar*; *nāmard* 'brute' < P *nāmard*.

   Some items ultimately of Arabic origin have been re-borrowed into Khuzestan Arabic from Persian, preserving the Persian meaning, e.g. KhA *bərɣi* 'electronic' < P *barɣ* 'electricity; lightning' < Arabic *barq* 'lightning'.

### 3.5.3 Verbs and adverbs

Khuzestan Arabic verbs and adverbs resulting from language contact are always morphologically integrated. These are either directly borrowed Persian verbs, e.g. *bannad* 'to close (e.g. the tap)' < P imperfect and present stem *band-* 'close';[24] *gayyər* 'to get stuck' < P *gīr kardan* 'to get stuck'; *ʕammər* 'to repair' < P *taʕmīr kardan* 'to repair'; *čassəb* 'to glue' < P *časb zadan* 'to glue'; *gəzar* 'to pass (time)' < P present stem *gozar-* 'to pass (time)' (see example (15) below)[25]; or Persian nouns

---

[24]Also common in the Gulf region and in Yemen (**BehnstedtWoidich2014**: 290).

[25]The verb *gəzar* is used only in phrases that refer to the "passing by" of life.

turned into Khuzestan Arabic (ad)verbs, e.g. *əb-zūr* 'by force' < P *zūr* 'power; violence; force'; and *zaḥəm* 'to bother s.o.' < P *zahmat* 'trouble, bother' (see examples (16) and (17) below).[26]

(15)  KhA, Aḥwāz (own data)
    čā  hāy    əl-ḥayāt lō  la?  təgzar              baʕad,      təmši.
    DP DEM.F DEF-life  or  no  pass.IMPF.3SG.F after_all go.IMPF.3SG.F
    'See, that is how life is, right? It passes by (quickly), it goes.'

(16)  KhA, Aḥwāz (own data)
    zaḥmīət-kum,              ʕafwan.[27]
    bother.PRF.1SG-2PL.M sorry
    'Sorry, I must have bothered you/caused you trouble.'

(17)  KhA, Aḥwāz (own data)
    mumkin azaḥm-ek              əb-šuɣla?
    possible  bother.IMPF.1SG-2SG.M with-issue
    'May I bother you with something/ask you a favour?'

### 3.5.4 Discourse elements

A range of Persian discourse elements have been borrowed by Khuzestan Arabic (cf. **MatrasShabibi2007**: 143–145),[28] e.g. KhA *ham*/*hamme* 'also, as well' < P *ham* and KhA *ham...ham* '(both)...and' < P *ham...ham*;[29] and KhA *hīč* 'nothing; no(t)... at all' < P *hīč*. [30]

---

[26]Khuzestan Arabic *zaḥme* is also used as a noun for a rebuke, e.g. *zaḥme ʕalīək!* 'Shame on you!', which would be expressed in a different way in Persian: *xeǧālat nemi keši?* 'Shame on you! (lit. 'Are you not ashamed?')'.

[27]A phrase often used when leaving after an invitation for dinner.

[28]Matras & **Shabibi2007** claim that the Persian conjunctions *agarče* and *bāīnke*, both meaning 'although/even though', and the Persian factual complementizer *ke* 'that' have also been borrowed by Khuzestan Arabic. However, I have found no evidence for their usage in my data.

[29]This discourse element is also known for Iraq (**Malaika1963**) and, like KhA *hast*/*hassət* 'there is' < P *hast* (**Ingham1973**: 25, fn.27), probably is an older borrowing.

[30]Shabibi (2006: 176–177) derives KhA *balkət* 'maybe, hopefully' from P *balke ham*, which can mean 'maybe' . A Turkish origin of this word seems more likely: cf. **Aksoy1963** for the existence of *belke*/*belkit* in Eastern Turkish dialects. **Malaika1963** also derives the Baghdadi Arabic *belki* 'rather, maybe' from Turkish, as does **Seeger2009** for *balki, balkīš, balkin* 'maybe; possibly; probably' in Ramallah Arabic.

The Khuzestan Arabic discourse elements *ḫō/ḫōš*[31] 'well; okay' < P *ḫō(b)/ḫōš* are often used phrase-initially. They are of Persian origin, but in the RL Khuzestan Arabic have partly adopted a different form and function. [32]

(18)  KhA, Aḥwāz (own data)
      ḫōš, š-ʕəd-na,         taʕay        əhna baba.
      DP   what-at-OBL.1PL come.IMP.SG.F here  father
      'Okay, what (else) do we have, come here, dear!'

Both *ḫō* and *ḫōš* are also often used in stories following the verb *gāl* 'to say'.

(19)  KhA, Fəllāḥīye/Aḥwāz (own data)
      lamman ɣada          mən ʕəd-hum,       gāl-la                  ḫō,
      when    leave.PRF.3SG.M from at-OBL.3PL.M say.PRF.3SG.M-3SG.M.DAT DP
      hāy       ər-rummānāt         š-asawwi          bī-hən?
      DEM.SG.F DEF-pomegranate.PL what-make.IMPF.1SG with-OBL.3PL.F
      'When he left them, he said to him, 'Well, what shall I do with these pomegranates?''

# 4  Conclusion

Because of the dominance of Persian in the Iranian educational system and work environment, the lack of influence from Modern Standard Arabic, and the long period of geographical proximity, the Persian-speaking society of southwest Iran has left many linguistic traces in the language of the Arabic-speaking community of Khuzestan.

Van Coetsem (2000: 59; cf. Lucas, 2015: 532) suggests that lexical, but not syntactic and phonological, transfer is to be expected under RL agentivity. However, Khuzestan Arabic phonology and syntax have been influenced by the SL Persian under RL agentivity, albeit to a much lesser extent than the lexicon.

Khuzestan Arabic does not show transfer of patterns from Persian in either inflectional or derivational morphology. However, we do find an adapted pattern replication of Persian phrasal verbs (with preservation of the Arabic word order).

---

[31]According to my informants and data, the form *ḫōb* is not used in Khuzestan Arabic (contrast **MatrasShabibi2007**: 143).

[32]In Persian, *ḫō(b)* is a discourse particle and simple adjective, and *ḫōš* just an adjective (see §**??**; **Shabibi2006**: 160). Thus the Persian adjective *ḫōš* has been desemanticized in Khuzestan Arabic to function as a discourse particle with the meaning 'well, okay' (**Shabibi2006**).

As to syntax and contact-induced word order changes, the alternative sentence construction with *čān* in sentence-final position can be explained as a result of Persian influence on Khuzestan Arabic. This change might have been triggered by the similar and very frequent Persian constructions with sentence-final *būdan*. Thus, we do have some syntactic change due to transfer under RL agentivity, which van Coetsem considered to be unexpected (see above).

Persian lexical items have often been borrowed in Khuzestan Arabic for novel concepts (lexical gaps), which is why semantic fields relating to technical or cultural innovations, education, and administration show the greatest amount of Persian borrowing. This also explains why nouns are generally more often transferred than verbs (cf. Lucas, 2015: 532). Persian words are regularly integrated into Khuzestan Arabic phonology and morphology, for example the Arabic internal plural is formed for Persian nouns. Also, many discourse particles have been transferred from Persian into Khuzestan Arabic. Some of them, e.g. *ham* 'also', had been in use generations ago among Arabic speakers in Khuzestan and beyond (Iraq, Gulf).

Of course, contact between Khuzestan Arabic and Persian has always been limited to certain social contexts (outside the family), especially for women, who had and still have much less access to education and employment and thus to the Persian-speaking world. This fact, and some structural differences between the languages, explain the limits of contact-induced language change in Khuzestan Arabic, especially in morphology and syntax.

Hopefully, future research on the dialects of Khuzestan will provide more empirical data on instances of contact-induced change. An enlarged database should especially provide further evidence concerning the development and extent of word order changes.

## Further reading

**Ingham2007**: Sketch grammar of Khuzestan Arabic.

**Ingham2005**: Discussion of Turkish and Persian borrowings in Khuzestan Arabic and North Eastern Arabian dialects.

**MatrasShabibi2007**: Article on contact-induced changes in Khuzestan Arabic based on **Shabibi2006**.

**Shabibi2006**: Unpublished dissertation on contact-induced changes in Khuzestan Arabic.

## Acknowledgements

## Abbreviations

ACC accusative
CONS construct state marker
COP copula
DAT dative
DEF definite
DEM demonstrative
DP discourse particle
EZ Persian Ezāfe
F feminine
IMP imperative
IMPF imperfect
INF infinitive
INDEF indefinite

KhA Khuzestan Arabic
M masculine
OBL oblique
P Persian
PL plural
PTCP participle
PRF perfect
PROG progressive
PRS present
PST past
REL relative particle
SG singular

```
%%move bib entries to  localbibliography.bib



@book{Ahadi2001,
address = {Wiesbaden},
author = {Ahadi, Shahram},
publisher = {Reichert},
title = {\textit{Verbergänzungen} \textit{und} \textit{zusammengesetzte} \textit{{\
year = {2001}
}
```

```
@book{Aksoy1963,
address = {Ankara},
booktitle = {\textit{Türkiyede} \textit{halk} \textit{ağzından} \textit{derleme}
editor = {Aksoy, Ömer A.},
publisher = {Türk Tarih Kurumu Basımevi},
sortname = {Aksoy, Omer A.},
title = {\textit{Türkiyede} \textit{halk} \textit{ağzından} \textit{derleme} \te
year = {1963}
}




@book{BehnstedtBehnstedt2014,
address = {Leiden},
author = {Behnstedt, Peter and Woidich, Manfred},
publisher = {Brill},
title = {\textit{Wortatlas} \textit{der} \textit{arabischen} \textit{{Dialekte}}
year = {2014}
}




@book{Brustad2000,
address = {\textit{The} \textit{syntax} \textit{of} \textit{spoken} \textit{Arab
author = {Brustad, Kristen E.},
publisher = {Georgetown University Press},
year = {2000}
}




@incollection{Daniel1986,
address = {London},
author = {Daniel, E. L.},
booktitle = {\textit{Encyclopaedia} \textit{Iranica}},
```

```
editor = {Yarshater, Ehsan},
pages = {210–214},
publisher = {Routledge \& Kegan Paul},
title = {Arab settlements in {Iran}},
volume = {2},
year = {1986}
}




@incollection{DePlanhol1986,
address = {London},
author = {De Planhol, X.},
booktitle = {\textit{Encyclopaedia} \textit{Iranica}},
editor = {Yarshater, Ehsan},
pages = {53–57},
publisher = {Routledge \& Kegan Paul},
title = {Ābādān, ii. {{M}}odern Ābādān},
volume = {2},
year = {1986}
}




@book{Denz1971,
address = {\textit{Die} \textit{Verbalsyntax} \textit{des} \textit{neuarabischen} \
author = {Denz, Adolf},
publisher = {Steiner},
year = {1971}
}




@book{Erwin1963,
address = {\textit{A} \textit{short} \textit{reference} \textit{grammar} \textit{o
```

```
author = {Erwin, Wallace M.},
publisher = {Georgetown University Press},
year = {1963}
}




@incollection{Gazsi2011,
address = {Berlin},
author = {Gazsi, Dénes},
booktitle = {\textit{The} \textit{{Semitic}} \textit{languages:} \textit{An} \te
editor = {Weninger, Stefan and Watson, Janet C. E. and Streck, Michael and Khan,
pages = {1015–2021},
publisher = {De Gruyter Mouton},
sortname = {Gazsi, Denes},
title = {{Arabic}-{Persian} language contact},
year = {2011}
}




@book{Grigore2007,
address = {Bucharest},
author = {Grigore, George},
publisher = {Editura Universitãţii din Bucureşti},
title = {\textit{L'arabe} \textit{parlé} \textit{à} \textit{Mardin:} \textit{Mon
year = {2007}
}




@book{Grotzfeld1965,
address = {Wiesbaden},
author = {Grotzfeld, Heinz},
publisher = {Harrassowitz},
```

```
title = {\textit{Syrisch-arabische} \textit{{Grammatik}} \textit{({Dialekt}} \text
year = {1965}
}




@book{Holes2016,
address = {\textit{Dialect,} \textit{culture,} \textit{and} \textit{society} \text
author = {Holes, Clive},
publisher = {Brill},
year = {2016}
}




@incollection{Ingham2007,
address = {Leiden},
author = {Ingham, Bruce},
booktitle = {\textit{Encyclopedia} \textit{of} \textit{{Arabic}} \textit{Language}
editor = {Versteegh, Kees},
pages = {571–578},
publisher = {Brill},
title = {Khuzestan {Arabic}},
year = {2007}
}




@incollection{Ingham2005,
address = {London},
author = {Ingham, Bruce},
booktitle = {\textit{Linguistic} \textit{convergence} \textit{and} \textit{areal} \
editor = {Csató, Éva Ágnes and Isaksson, Bo and Jahani, Carina},
pages = {173–180},
publisher = {Routledge Curzon},
```

title = {{Persian} and {Turkish} loans in the {Arabic} dialects of {North} {East
year = {2005}
}


@incollection{Ingham1991,
address = {Wiesbaden},
author = {Ingham, Bruce},
booktitle = {\textit{{Semitic}} \textit{studies:} \textit{In} \textit{honor} \te
editor = {Kaye, Alan S.},
pages = {714–728},
publisher = {Harrassowitz},
title = {Sentence structure in Khuzistani {Arabic}},
year = {1991}
}


@article{Ingham1976,
author = {Ingham, Bruce},
journal = {\textit{Bulletin} \textit{of} \textit{the} \textit{School} \textit{of
pages = {62–82},
title = {Regional and social factors in the dialect geography of southern {Iraq}
volume = {39},
year = {1976}
}


@article{Ingham1973,
author = {Ingham, Bruce},
journal = {\textit{Bulletin} \textit{of} \textit{the} \textit{School} \textit{of
pages = {523–553},
title = {Urban and rural {Arabic} in Khūzistān},

```
volume = {36},
year = {1973}
}




@incollection{Lucas2015,
address = {New York},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{{Routledge}} \textit{handbook} \textit{of} \text
editor = {Bowern, Claire and Bethwyn, Evans},
pages = {519–536},
publisher = {Routledge},
title = {Contact-induced language change},
year = {2015}
}




@article{Lucas2012,
author = {Lucas, Christopher},
journal = {\textit{Diachronica}},
pages = {275–300},
title = {Contact-induced grammatical change: {{T}}owards an explicit account},
volume = {29},
year = {2012}
}




@book{Majidi1990,
address = {\textit{Strukturelle} \textit{Grammatik} \textit{des} \textit{Neupersis
author = {Majidi, Mohammad-Reza},
publisher = {Buske},
year = {1990}
```

}

@book{Majidi1986,
address = {\textit{Strukturelle} \textit{Grammatik} \textit{des} \textit{Neupers
author = {Majidi, Mohammad-Reza},
publisher = {Buske},
year = {1986}
}

@book{Malaika1963,
address = {\textit{Grundzüge} \textit{der} \textit{Grammatik} \textit{des} \text
author = {Malaika, Nisar},
publisher = {Harrassowitz},
year = {1963}
}

@incollection{MatrasMatras2007,
address = {Berlin},
author = {Matras, Yaron and Shabibi, Maryam},
booktitle = {\textit{Grammatical} \textit{borrowing} \textit{in} \textit{cross-l
editor = {Matras, Yaron and Bossong, Georg and Comrie, Bernard},
pages = {137–149},
publisher = {Mouton de Gruyter},
title = {Grammatical borrowing in Khuzistani {Arabic}},
year = {2007}
}

```
@incollection{Oberling1986,
address = {London},
author = {Oberling, P.},
booktitle = {\textit{Encyclopaedia} \textit{Iranica}},
editor = {Yarshater, Ehsan},
pages = {215–219},
publisher = {Routledge \& Kegan Paul},
title = {Arab tribes of {Iran}},
volume = {2},
year = {1986}
}




@incollection{Pat-El2017,
address = {Leiden},
author = {Pat-El, Na'ama},
booktitle = {\textit{{{Arabic}}} \textit{in} \textit{context:} \textit{Celebrating}
editor = {Al-Jallad, Ahmad},
pages = {441–475},
publisher = {Brill},
sortname = {Pat-El, Na'ama},
title = {Digging up archaic features: {{N}}eo-{{Arabic}} and comparative Semitics
year = {2017}
}




Procházka, Sthephan. 2018. The Northern Fertile Crescent. In Holes, Clive (ed.), \t




@incollection{Ratcliffe2005,
address = {London},
author = {Ratcliffe, Robert R.},
```

booktitle = {\textit{Linguistic} \textit{convergence} \textit{and} \textit{areal
editor = {Csató, Éva Ágnes and Isaksson, Bo and Jahani, Carina},
pages = {141–159},
publisher = {Routledge Curzon},
title = {Bukhara {{Arabic}}: {{A}} metatypized dialect of {{Arabic}} in {Central
year = {2005}
}




@incollection{Retsö2009,
address = {Oxford},
author = {Retsö, Jan},
booktitle = {\textit{Relative} \textit{clauses} \textit{and} \textit{genitive} \
editor = {Watson, Janet and Retsö, Jan},
pages = {3–33},
publisher = {Oxford University Press},
sortname = {Retso, Jan},
title = {Nominal attribution in {Semitic}: {{T}}ypology and diachrony},
year = {2009}
}




@incollection{Sakel2007,
address = {Berlin},
author = {Sakel, Jeanette},
booktitle = {\textit{Grammatical} \textit{borrowing} \textit{in} \textit{cross-l
editor = {Matras, Yaron and Bossong, Georg and Comrie, Bernard},
pages = {15–29},
publisher = {Mouton de Gruyter},
title = {Types of loan: {{M}}atter and pattern},
year = {2007}
}

```
@book{Seeger2009,
address = {\textit{Der} \textit{arabische} \textit{Dialekt} \textit{der} \textit{D
author = {Seeger, Ulrich},
publisher = {Harrassowitz},
year = {2009}
}




@misc{Shabibi2006,
author = {Shabibi, Maryam},
note = {} University of Manchester. (Unpublished doctoral dissertation).},
title = {\textit{Contact-induced} \textit{grammatical} \textit{changes} \textit{in
year = {2006}
}




@book{VanCoetsem2000,
address = {Heidelberg},
author = {Van Coetsem, Frans},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theory}
year = {2000}
}




@incollection{Versteegh2010,
address = {Oxford},
author = {Versteegh, Kees},
booktitle = {\textit{The} \textit{handbook} \textit{of} \textit{language} \textit{
editor = {Hickey, Raymond},
```

```
pages = {634–651},
publisher = {Wiley-Blackwell},
title = {Contact and the development of {Arabic}},
year = {2010}
}




@book{Winford2003,
address = {\textit{An} \textit{introduction} \textit{to} \textit{contact} \texti
author = {Winford, Donald},
publisher = {Blackwell},
year = {2003}
}




@book{WoodheadWoodhead1967,
address = {Washington D.C.},
author = {Woodhead, D. R. and Beene, Wayne},
note = {Press.},
publisher = {Georgetown Univ},
title = {\textit{A} \textit{dictionary} \textit{of} \textit{Iraqi} \textit{{{Ara
year = {1967}
}
```

# Chapter 3

# Contact-induced change in Maghrebi Arabic

Adam Benkato

University of California, Berkeley

This chapter gives an overview of contact-induced changes in the Maghrebi dialect group in North Africa. It includes both a general summary of relevant research on the topic and a selection of case studies which exemplify contact-induced changes in the areas of phonology, morphology, syntax, and lexicon.

## 1 The Maghrebi Arabic varieties

In Arabic dialectology, Maghrebi is generally considered to be one of the main dialect groups of Arabic, denoting the dialects spoken in a region stretching from the Nile delta to Africa's Atlantic coast – in other words, the dialects of Mauritania, Morocco, Algeria, Tunisia, Libya, parts of western Egypt, and Malta. The main isogloss distinguishing Maghrebi dialects from non-Maghrebi dialects is the first person of the imperfect, as shown in Table 1 (cf. Lucas & Čéplö this volume).[1]

This Maghrebi group of dialects is in turn traditionally held to consist of two subtypes: those spoken by sedentary populations in the old urban centers of North Africa, and those spoken by nomadic populations. The former of these, usually referred to as "pre-Hilali" (better: "first-layer") would have originated with the earliest Arab communities established across North Africa (~7th to 8th centuries CE) up to the Iberian peninsula. The latter of these, usually referred to as "Hilali" (better: "second-layer"), is held to have originated with the westward migration of a large group of Bedouin tribes (~11th century CE) out of the Arabian

---

[1]More about the exact distribution of this isogloss can be found in **Behnstedt2016**.

Table 1: First person imperfect 'write' in Maghrebi and non-Maghrebi Arabic

|  | Non-Maghrebi | | Maghrebi | |
|  | Classical Arabic | Baghdad Arabic | Casablanca Arabic | Maltese |
| --- | --- | --- | --- | --- |
| Singular | *aktub* | *aktib* | *nəktəb* | *nikteb* |
| Plural | *naktub* | *niktib* | *nkətbu* | *niktbu* |

peninsula and into North Africa via Egypt. Their distribution is roughly as fol-lows.[2] First-layer dialects exist in cities such as Tunis, Kairouan, Mahdia, Sousse, Sfax (Tunisia), Jijel, Algiers, Cherchell, Tlemcen (Algeria), Tangier, Tetuan, south-ern Rif villages, Rabat, Fez, Taza, so-called "northern" dialects (Morocco), Mal-tese, and formerly Andalusi and Sicilian dialects; most Judeo-Arabic dialects for-merly spoken in parts of North Africa are also part of this group. Second-layer di-alects are spoken by populations of nearly all other regions, from western Egypt, through all urban and rural parts of Libya, to the remaining urban and rural parts of Algeria and Morocco. Though some differences between these two subtypes are clear (such as [q, ʔ, k] vs. [g] for *q), there have probably been varying lev-els of interdialectal mixture and contact since the 11th century CE. In many cases, first-layer varieties of urban centers have been influenced by neighboring second-layer ones, leading to new dialects formed on the basis of inter-dialectal contact. It is important to note that North Africa is becoming increasingly urbanized and so not only is the traditional 'sedentary-nomadic' distinction anachronistic (if it was ever completely true), but also that intensifying dialect contact accompany-ing urbanization means that new ways of thinking about Maghrebi dialects are necessary. It is also possible to speak of the recent but ongoing koinéization of multiple local varieties into supralocal or even roughly national varieties—thus one can speak, in a general way, of "Libyan Arabic" or "Moroccan Arabic". This chapter will not deal with contact between mutually intelligible varieties of a language although this is equally important for the understanding of both the history and present of Maghrebi dialects.[3]

---

[2]More will not be said about the subgroups of Maghrebi dialects that have been proposed. For more details about the features and distribution of Maghrebi dialects see **Pereira2011**; for more detail on the complex distribution of varieties in Morocco see **Heath2002**.

[3]The emergence of new Maghrebi varieties resulting from migration and mixture is discussed in **Pereira2007** and **Gibson2002**, for example. The oft-cited distinction between urban and nomadic dialects is also problematized by the existence of the so-called rural or village dialects (though this is also a problematic ecolinguistic term), on which see **Mion2015**. Dialect contact outside of the

## 2  Languages in contact

Contact between Arabic and other languages in North Africa began in the late 7th-century CE, when Arab armies began to spread westward through North Africa, reaching the Iberian peninsula by the early 8th century CE and founding or occupying settlements along the way. Their dialects would have come into contact with the languages spoken in coastal regions at that time, including varieties of Berber and Late Latin, and possibly even late forms of Punic and Greek. The numbers of Arabic-speakers moving into North Africa at the time of initial conquests was likely to have been quite small.[4] By the time of the migration of Bedouin groups beginning in the 11th century, it is doubtful that languages other than Berber and Arabic survived in the Maghreb. The Arabization of coastal hinterlands and the Sahara increased in pace after the 11th century. Berber varieties continue to be spoken natively by millions in Morocco and Algeria, and by smaller communities in Libya, Tunisia, Mauritania, and Egypt. Any changes in an Arabic variety due to Berber are almost certainly the result of Berber speakers adopting Arabic rather than Arabic speakers adopting Berber – the sociolinguistic situation in North Africa is such that L1 Arabic speakers rarely acquire Berber.

Beginning in the 16th century, most of North Africa came under the control of the Ottoman Empire and thus into contact with varieties of Turkish, although the effect of Turkish is essentially limited to cultural borrowings (see below, 3.4). The sociolinguistic conditions in which Turkish was spoken in North Africa are poorly understood.

The advent of colonialism imposed different European languages on the region, most prominently French (in Mauritania, Morocco, Algeria, and Tunisia), Italian (in Libya), and Spanish (in Morocco). Romance words in dialects outside of Morocco may also derive from forms of Spanish (via Andalusi refugees to North Africa in the 16th/17th centuries) or from the Mediterranean Lingua Franca.[5]

The effects on Maghrebi Arabic of contact with Chadic (Hausa) or Nilo-Saharan (Songhay, Tebu) languages is largely unstudied since in most cases data from the relevant Arabic varieties is lacking. Yet some borrowings from these languages can be found in Arabic and Berber varieties throughout the region (**Souag2013**).[6] Lastly, Hebrew loans are present in most Jewish Arabic dialects of North Africa (**Yoda2013**), though unfortunately these dialects hardly exist anymore.

To restate these facts in Van Coetsem's (**Coetsem1988**; **Coetsem2000**) terms,

---

Maghreb is discussed by Cotter (this volume).

[4]See Heath (this volume) for discussion of Late Latin influence in Moroccan Arabic dialects.

[5]On the Lingua Franca see Nolan (this volume).

[6]See also **Souag2016** for an overview of contact in the Sahara region not limited to Arabic.

there are two major contact situations at work in Maghrebi Arabic in general, though the specifics will of course differ from variety to variety. The first is change in Arabic driven by source-language (Berber) dominant speakers; this transfer type is IMPOSITION. The second is change in Arabic driven by recipient-language (Arabic) dominant speakers where the source language is a European colonial language; this transfer type is called BORROWING.[7] So far, 'dominance' describes linguistic dominance, that is, the fact that a speaker is more proficient in one of the languages involved in the contact situation. However, social dominance, referring to the social and political status of a language (**Coetsem1988**: 13), is also important, especially in North Africa.

## 3 Contact-induced changes in Maghrebi dialects

### 3.1 Phonology

Changes in Maghrebi Arabic phonology due to contact with Berber are difficult to prove. There are several cases, for example, where historical changes in Arabic phonology may be argued to be the result of contact with Berber *or* the result of internal developments. These include the change of *ğ to /ž/ in many varieties, or the emergence of phonemic /ẓ/ (**Souag2016**). Another example, the pronunciation /ṭ/ in some first-layer varieties where most Arabic varieties have /ḍ/ has also been explained as a result of Berber influence, or as unclear directionality (**Kossmann2013book**), while **Al-Jallad2015** argues that it is actually an archaism within Arabic.

The merger in Arabic of the vowels *a and *i (and even *u) to a single phoneme /ə/ in some, especially first-layer, varieties, is often attributed to Berber influence, as many Berber varieties have only a single short vowel phoneme /ə/. However Kossmann (**Kossmann2013book**: 171–174) points out that Berber also merged older *ă and *ə to a single phoneme /ə/ and that it cannot be proven that the reduction happened in Berber before it happened in Arabic. Hence, again the directionality of influence is difficult to show.

Related to this development is also that many Maghrebi varieties disallow vowels in light syllables (often described as the deletion of short vowels in open syllables), such that *katab 'he wrote' > Tripoli ktəb or *kitāb 'book' > Algerian ktāb.[8]

---

[7]Another good illustration of the two transfer types in the Van Coetsemian framework can be found in Winford (**Winford2005**: 378–381).

[8]Since the short vowels merge to schwa in many Moroccan and Algerian varieties, vowel length is no longer contrastive and it is common to transcribe e.g. *ktab* rather than *ktāb*.

Meanwhile, second-layer varieties often do allow vowels in light syllables (e.g. Benghazi *kitab* 'he wrote', Douz *m^i šē* 'he went'). While proto-Berber and some modern varieties allow vowels in light syllables, most Berber varieties of Algeria and Morocco do not. This is another example of a similar development wherein the directionality of influence is unclear (see Souag **Souag2017**: 62–65 for further discussion).

In the Arabic variety of Ghomara, northwest Morocco, *d and *t are spirantized to /ð/ and /θ/ initially (*d only), postvocalically and finally (**Naciri-Azzouz2016**): e.g. *māθəθ* 'she died' (*mātat), *warθ* 'inheritance' (although etymologically *warθ, dialects of the wider Jbala region of Morocco have no interdentals so *wart), *ðāba* 'now' (*dāba), *ḫəðma* 'work' (*ḫidma), *wāḥəð* 'one' (*wāḥid). Naciri-Azzouz points out that the distribution of spirantization is the same as in Ghomara Berber, a variety spoken by groups in the same region.[9]

New phonemes have been borrowed into Maghrebi varieties through contact with European languages: for example, /p/ and nasalized vowels in more recent French loans in Tunisian Arabic, or /v, č, ǧ/ in Italian loans in Libyan Arabic (*grīǧū* 'gray' < *grigio*).

## 3.2 Morphology

In the realm of morphology, changes in Arabic varieties due to contact vary depending on whether the relationship between Arabic and the contact language is substratal, adstratal, or superstratal.

Morphological influence from Berber on Arabic varieties of the northern Maghreb is not overly common.[10] In some places where Berber/Arabic bilingualism is or was more common, contact has led to the borrowing of Berber nouns into Arabic together with their morphology, a phenomenon known as "parallel system borrowing".[11] In Ḥassāniyya, for example, many nouns have been transferred together with their gender and number marking.[12] In the dialect of Jijel, Berber singular nouns are transferred together with their prefixes (*āwtūl* 'hare', cf. Kabyle

---

[9]The Berber variety of Ghomara exhibits an extreme amount of influence from dialectal Arabic, see Mourigh (**Mourigh2015**). Kossmann (**Kossmann2013book**) writes that given the existence of parallel morphological systems for virtually all grammatical categories (nominal, adjectival, pronominal and verbal morphology) and high loanword count (more than 30% of basic lexicon is Arabic) it would be possible to call Ghomara Berber a mixed language.

[10]Documentation of the varieties where such influence would be more expected, such as Arabic-speaking towns in the otherwise Berber-speaking Nafusa Mountains in Libya, is lacking.

[11]For a closer look at parallel system borrowing in the context of Arabic and Berber contact, see **Kossmann2010**, mostly discussing the borrowing of Arabic paradigms into Berber.

[12]See Taine-Cheikh (this volume).

*āwtūl*); plurals are then formed in a way which resembles Berber but is not exact (Jijel *āsrəf, āsərfən* 'bushe(s)', cf. Kabyle Berber *āsrəf, īsərfən*); moreover, the prefix *ā-* is also used with nouns of Arabic origin (*āfḫəd* 'thigh', Arabic \**faḫaḍ*) (**Marçais1956**).

In Algeria and Morocco the circumfix *tā-...-t*, which occurs on feminine nouns in Berber, can derive abstract nouns (e.g. Jijel *tākəbūrt* 'boasting', *tāwəḥḥūnt* 'having labor pains') and in Moroccan Arabic *tā-...-t* is the regular way of forming nouns of professions and traits (e.g. *tānəžžāṛt* 'carpentry') (**Kossmann2013chapter**).

The verbal morphology of Arabic dialects is much less affected by Berber, though Ḥassaniya again provides an interesting example. It has a causative prefix *sä-* used with both inherited Arabic verbs and borrowed Berber verbs, and most likely to be borrowed from Berber causative forms in *s-/š-* (**Taine-Cheikh2008**).

Turkish influence on morphology is restricted to the suffix *-ği/-ži* (< *-ci*) used to indicate professions and borrowed widely into Arabic dialects in general. In Tunisia, its use has been extended to derive adjectives of quality from nouns (*sukkārži* 'drunkard') and has also even been added to borrowed French nouns (*bankāži* 'banker' < French *banque*). As **Manfredi2018** points out, the productivity of this borrowed derivational morpheme is one example of how recipient-language agentivity can introduce morphological innovations based on borrowing.

French (and other Romance) verbs are also routinely borrowed into Maghrebi varieties. **Talmoudi1986** discusses their integration into different forms of the verbal system of Tunisian Arabic, e.g. *mannak* < *manquer* or *(t)rānā* < *trainer*.

## 3.3 Syntax

Syntax is often the least documented aspect of the grammar of Maghrebi Arabic varieties and research on contact-induced changes in syntax is still in its infancy. Much attention has been devoted recently to explaining the rise of bipartite negation in Arabic and Berber; in varieties of both languages the word for 'thing' (Arabic *šayʔ*, Berber \**k̆ara*) has been grammaticalized postverbally in a marker of negation:

(1)   Arabic (Benghazi)
   **mā**-šift-hā-**š**
   NEG-see.PRF.1SG-3SG.F-NEG
   'I didn't see her'.

(2)   Berber (Tarifit)

**wā**  t-ẓṛiɣ        **ša**
NEG  3SG.F-see.PFV.1SG  NEG
'I didn't see her'.

Although some accounts give no attention to Berber, while others attribute the Arabic development solely to Berber, the development in both languages in the same contexts is probably not a coincidence; though there is no current consensus on the direction of transfer – see Lucas (this volume) for discussion.[13] However, it must be noted that not at all Berber varieties have double negation (e.g. Tashelhiyt *ur nniɣ ak* 'I didn't tell you' where the only negator is *ur*).

In another area, recent work on the variety of Tunis has yielded interesting conclusions: while possessives with French nouns are overwhemingly analytic (*l-prononciation mtēʕ-ha* 'her pronunciation') and those with Arabic nouns are almost as overwhelmingly synthetic (*nuṭq-u* 'his pronunciation'), the frequent occurence of French loan nouns may be triggering an increase in the overall frequency of analytical possessives over syntactic ones, including those with Arabic nouns (**Sayahi2015**).

The remainder of this section will discuss one particularly interesting case: the first-layer dialect of Jijel, a city in eastern Algeria. At the time of its description (**Marçais1956**), it showed little influence from second-layer varieties, but displayed wide-ranging influence from Berber in multiple domains. In a recent article, **Kossmann2014** has demonstrated how a Berber marker of non-verbal predication was adopted into the Arabic dialect of Jijel as a focus marker. Here I will briefly summarize Kossmann's arguments with a few examples. In the Jijel dialect, as described by Marçais and reanalyzed by Kossmann, a morpheme *d* occurs in the following syntactic contexts (examples (3)–(7) are all from **Kossmann2014**, who retranscribes from Marçais' texts): before non-verbal predicates (3), in clefts with a noun/pronoun in the cleft (4), in secondary predication with a specific noun (5), as a marker of subject (or object) focus (6), and in left-moved focalizations (7).

---

[13]See **Lucas2007**; **Lucas2010**; **Lucas2018** and **Souag2018** for further discussion of the grammaticalization of 'thing' for indefinite quantification and polar question marking in Arabic and Berber. **Kossmann2013book** surveys the situation in the Berber languages. See **Lafkioui2013** for an overview of negation in especially Moroccan Arabic, as well as discussion of a variety of Moroccan Arabic which features the discontinuous morpheme *mā- ... -bū*, where the latter part has been borrowed from Tarifit.

(3)  l-lila       d-əl-ʕid
     DEF-night D-DEF-feast
     'Tonight is the feast.'

(4)  d-hum   əddə šraw-əh         qbəl-ma       nəzdad
     D-3PL.M REL   buy.PRF.3PL.M-3SG.M before-COMPAR be.born.IMPF.1SG
     'It is them who bought it before I was born.'

(5)  ṛa-na    nqəṭṭʕu-č      d əṭ-ṭraf
     PRZ-1PL cut.IMPF.1PL-SG D DEF-pieces
     'We will cut you (into) pieces.'

(6)  tkəṣṣṛət        d l-idura
     break.PRF.3SG.F D DEF-bowl
     'The bowl has broken.'

(7)  qalu      d əṛ-ṛbiʕ      dəḫlət
     say.PRF.3PL D DEF-spring enter.PRF.3SG.F
     'They say spring has come.'

Although previous analyses attempted to explain *d* within Arabic, Kossmann notes that an Arabic-internal derivation of *d* is impossible. However, Kabyle, the Berber language neighboring the Jijel area has an element *d* (realized [ð] due to spirantization in Kabyle) which is used in (pro)nominal predicates (8), cleft constructions (9), and secondary predication when non-verbal (10). Examples (8)–(10) are all Kabyle Berber, taken from **Kossmann2014**. This element *d* is attested in Berber more widely, too, and is likely reconstructible to older stages of the language.

(8)  d-yəlli-m
     D-daughter-2SG.F
     'Is it your daughter?'

(9)  d-ay-ən  i  d-tənna          abrid amənzu
     D-this-DEIC REL hither-say.3SG.F road  first
     'This is what she said the first time.'

(10) ad   nəǧʕəl  iman nn-əɣ  d-inəbgiwən n  ṛəbbi
     MOD make.1PL self  GEN-1PL D-guests   GEN lord
     'We shall pretend to be beggars (lit. guests of God).'

Thus Berber *d* is the best candidate for the origin of Jijel Arabic *d*, though its usage in (Kabyle) Berber (where it is a primarily marker of syntactic organization) differs from that of Jijel Arabic (where it is mainly a marker of information structure). In a simplified scenario with a Berber variety as source language and Jijel Arabic as recipient, *d* would likely have been imposed into Jijel Arabic with its exact Berber functions. As Kossmann notes, though, speech communities are full of variation and language contact is a "negotiation between the frequency of non-native speech and the prestige of the native way of speaking" (**Kossmann2014**). Kossmann thus proposes a scenario in which larger groups of Berber speakers switched to a variety of Jijel Arabic and began imposing their own *d*; the native Jijel Arabic speakers, fewer in number, began adopting *d* but understood it differently and interpreted it as a focus marker, introducing it into new contexts; eventually the variety of Jijel Arabic with *d* in all these functions became nativized. Per **Kossmann2014**, two processes would have taken place: the transfer of a source-language feature by speakers dominant in the source language (Berber), followed by the borrowing of this feature by speakers dominant in the recipient language (Arabic), and its eventual regularization in that variety. Jijel Arabic is an excellent example of what may happen when large numbers of Berber speakers switch to Arabic.

## 3.4 Lexicon

Much work on contact and Maghrebi Arabic has focused on loanwords, the most salient effects of borrowing, with secondary attention to their phonological or morphological adaptation. The concept of social dominance has particular relevance for borrowing: in the North African context, the colonial languages, especially French, have high social status for both Arabic and Berber native speakers. One also must modify the idea of linguistic dominance to include those who acquire two languages natively (2L1 speakers; see (**Lucas2015**)), definitely the case for certain speakers of Berber and Arabic in North Africa.

Unsurprisingly, we see firstly that the majority of words borrowed into Arabic varieties are nouns, and secondly that the lexical domains into which these borrowings fall are often restricted. Social dominance seems to play role in the nature of the nouns borrowed.

Berber loans are found in most Maghrebi varieties, though their number ranges from only a handful of words in the east to many more in the west (cf. §3.2 above). Almost all Maghrebi varieties borrow the words *ž(i)ṛāna* 'frog' and *fakrūna* 'turtle', while in some oases Berber influence in agricultural terminology can be seen. Again, the documentation of the relevant varieties is often insufficient.

Several studies on contact between Maghrebi Arabic varieties and European languages exist. For French in Morocco, **Heath1989** argues that code-switching and borrowing are essentially the same in a bilingual community which has established borrowing routines.[14] For French in Tunisia, **Talmoudi1986** analyzes the phonological and morphological adaptation of French verbs into Arabic. **Sayahi2014** gives a broader view of lexical borrowing in diglossic/bilingual communities, focusing on French in Tunisia and Spanish in Morocco. Italian in Tunisia is studied briefly by **Cifoletti1994**. Studies of contact with Turkish are limited to discussion of loanwords: on Morocco see **Procházka2012**; on Algeria, see **BenCheneb1922**, to be read with review by G.S. Colin (**Colin1999**).

The remainder of this section will consider the influence of Turkish and Italian on Libyan Arabic (henceforth LA), a hitherto underresearched topic. Uniquely in the Maghreb region there is at present no superstratum language spoken widely by Arabic speakers in Libya, while there are also fewer Berber speakers than in Algeria or Morocco. As far as documented varieties of LA (Tripoli and Benghazi) go, contact situations are historical and not active.

There seems to be an impression among dialectologists that LA varieties have the largest number of Turkish loans, though there is not a published basis for this. **Procházka2005** suggests that the number of (Ottoman) Turkish loans in a given Arabic dialect is proportional to the length and intensity of Ottoman rule. By this criteria Libya should have quite a few, as the regions now constituting Libya were under control of the Ottoman Empire from 1551 to 1911, but Procházka estimates that the dialect would show 200 to 500 surviving loans, less than in other dialects. Another important factor is likely to be that Libya's population was very small during the period of Ottoman rule so that the long-term presence of even a few thousand Turkish speakers could have had a significant effect. However, I cannot yet offer a statistical analysis of Turkish words in Libyan Arabic.[15] It is clear so far, though, that the effects of Turkish on LA can mainly be seen in

---

[14]**Coetsem1988** notes that for bilingual speakers who have a balance in linguistic dominance between the two languages, the separation between the two transfer types (borrowing and imposition) will be weaker. Hence, either of the two dominant languages can serve as the recipient language in codeswitching behavior. **Winford2005** (2005, esp. 394–396), expanding on Van Coetsem's framework, points out that code-switching is inherently linked to the borrowing transfer type. In the Maghreb, this scenario is possible for Berber–Arabic bilinguals as well as for some French–Arabic bilinguals.

[15]The only study dedicated to Turkish loans in LA is **Türkmen1988**, who lists 90 words. However, the basis for his wordlist seems unclear and several items are either spurious or incorrect (e.g. there is no word *kabak* 'pumpkin' in Benghazi Arabic but there is *bkaywa* 'pumpkin', identified by **Souag2013** as a loan from Hausa). Turkish words in Libyan Arabic cited here are from the Benghazi variety, author's data.

the lexicon and, in my data, almost entirely in nouns. In terms of their semantic domains, **Procházka2005** points out that the majority of Turkish loans in Arabic dialects in general fall into three categories, roughly described as: private life; law, government, social classes; and army, war. By far the majority of surviving loans would belong to the first of these classes (such as *šīšma* 'tap' < *çeşme*, *dizdān* 'wallet' < *cüzdan*), or second (such as *fayramān* 'order' < *ferman*, *ḥafḏ̣a* 'week' < *hafte*,) while I suspect that words from the third class are increasingly rarer. Outside of these, only a few words other than nouns seem to be present, such as *duġri* 'straight ahead' and *balki* 'maybe'. The time passed since Turkish was last actively spoken in Libya no doubt means that the number of Turkish loans actively used by speakers has been decreasing.

LA is unique among Maghrebi varieties in having had Italian as the main European contact language. Italian had a presence in what is now Libya from the 1800s, but this was mainly limited to the Tripolitanian Jewish community and wealthy merchant families. The Italian colonization of Libya officially began in 1911; though the majority of the region was not brought under Italian control until the early 1930s, large numbers of Italian colonists had begun to settle in Libya in the 1920s. From that period until 1970, when remaining Italian citizens were expelled from the country, Italians made up 15% or more of the population and the language was in widespread use. From the 1970s on, Italian was scarcely used in Libya, and the teaching of foreign languages was banned in 1984, not to return again until 2005.[16] Many of the postwar generation spoke (and still speaks) Italian, though they rarely use it anymore, but few Libyans of younger generations do. The 1920s to the 1970s can thus be regarded as the main period of contact between LA and Italian.[17] However, the concentration of Italians differed from region to region and thus may have influenced local varieties differently. The primary study devoted to analyzing Italian loans in LA is that of **Abdu1988** who, focusing on the variety of Tripoli, draws up a list of nearly 700 items (a few are misidentified), of which about 50% were recognized by a majority of those surveyed. Some 93% of these are nouns and the remainder are practically all derived from nouns or adjectives, such as *bwōno* 'well done!' < *buono* 'good' or *faryaz* 'to go out of order' < Italian *fuori uso*.[18] Abdu's study **Abdu1988** groups Italian

---

[16]For more information on the return of Italian instruction to Libya, see **DAnna2018**.

[17]The Italian words in Yoda's study of Tripoli Judeo-Arabic (**Yoda2005**) need to be seen slightly differently than Italian words in non-Jewish dialects, owing to a different history of the Tripolitanian Jewish community with Italy.

[18]See **Abdu1988** and **DAnna2018**. Some denominal verbs are cited by Abdu, but more extensive data might reveal several more in use: for example in the variety of Benghazi, I identified *fuṛan* 'to brake (intransitive)' < *frayno* 'brake' < Italian *freno*, not listed by Abdu.

loans into some 22 semantic categories, the vast majority of which relate to material culture. Examples of these from the Benghazi variety are *byāmbu* 'lead' < *piombo*, *bōskō* 'zoo' < *bosco*, *furkayta* 'fork' < *forchetta*, *maršabīdi* 'sidewalk' < *marciapiede* (author's data).

As **DAnna2018** points out, the adaptation of Italian words to Libyan Arabic phonology varies: new phonemes, particularly [v] and [č], sometimes occur but are sometimes adapted to the dialects' pre-existing phonologies, an indication of "subsidiary phonological borrowing" (**Coetsem1988**: 98). Of course, the maintenance of new phonemes often depends on speakers continuing to have access to the source language; as this is no longer the case in Libya, Italian borrowings in Libyan Arabic are traversing a different trajectory than French borrowings in other Maghrebi varieties, where only the oldest borrowings have been phonologically integrated.

The overwhelming majority of surviving Turkish and Italian loans in Libyan Arabic are nouns, widely acknowledged to be the most easily-borrowed word class due to their being the least disruptive of the recipient language's argument structure (**Myers-Scotton2002**), though a few verbs derived dialect-internally do exist. Furthermore, almost all the nouns are cultural borrowings — "lexical content-words that denote an object or concept hitherto unfamiliar to the receiving society, terminology related to institutions that are the property of the neighboring [or colonizing] culture, and so on" (**Matras2011**). Cultural borrowings are to be differentiated from core borrowings, the latter being words that more or less duplicate already existing words and which originate in a bilingual codeswitching context. These facts lead us to conclude that Turkish and Italian borrowings in Libyan varieties would be from (1) to (2) on the borrowing scale proposed by **Thomason1988**. While (1) of the scale involves lexical borrowing of non-basic vocabulary only, (2) includes some function words as well as new phones appearing in those loanwords. Colonial language contact situations are typically ones of recipient-language agentivity, as the number of indigenous people learning the colonial language is many times more than the number of colonizers learning indigenous languages. Without a longer period of sustained bilingualism or language education motivated by continued contact with the metropole, Italian has affected Libyan Arabic to a much less degree than French has Libya's Maghrebi neighbours.

# 4 Conclusion

The general parameters of the Maghrebi linguistic landscape and contact situations are relatively well understood. However, more documentation of Maghrebi varieties is needed, and more specifically, of those where contact situations – especially with Berber – may have existed. Additionally, further research into the sociolinguistic factors affecting bilingualism in Berber and Arabic, or regarding the intersection of diglossia with bilingualism, will no doubt add to our knowledge of the parameters of contact-induced change more generally. Finally, inter-dialectal contact as well as the gradual rise of national or at least supra-local varieties certainly merits continuing attention.

# Further reading

**Kossmann2013book** is the most extensive study so far of Berber–Arabic contact, written from a Berberological point of view but important for Arabists.
**Sayahi2014** studies the intersection of dialects, Standard Arabic, and French/Spanish in Tunisia and Morocco.
**Souag2016** summarizes contact in the Saharan region among Arabic, Berber, Hausa, Songhay, Chadic, etc.

# Abbreviations

| | |
|---|---|
| COMPAR | comparative |
| DEF | definite article |
| DEIC | deictic particle |
| F | feminine |
| GEN | genitive marker |
| IMPF | imperfect (prefixal conjugation) |
| M | masculine |
| MOD | modal particle |
| NEG | negative |
| PL | plural |
| PFV | perfective |
| PRF | perfect (suffix conjugation) |
| PRZ | presentative |
| REL | relative |
| SG | singular |

*Adam Benkato*

## Acknowledgements

# Chapter 4

# Nigerian Arabic

Jonathan Owens

University of Bayreuth

Nigerian Arabic displays an interesting interplay of maintenance of inherited structures along with striking contact-induced innovations in a number of domains. This article summarizes the various domains where contact-based change has occurred, concentrating on those less studied not only in Arabic linguistics, but in Linguistics in general, namely idiomatic structure and an expanded functionalization of demonstratives. Methodologically, comparative corpora are employed to demonstrate the degree of contact-based influence.

## 1 1. Historical and linguistic background

Nigerian Arabic (NA) is spoken by perhaps – there are no reliable demographic figures from the last 50 years – 500,000 speakers. These are found mainly in northeast Nigeria in the state of Borno where their homeland is concentrated along the Cameroon-Chad border as far south as Banki, spreading westwards towards Gubio (see Map), and south of Maiduguri towards Damboa. Mirroring a larger trend in Nigerian demographics, the past 40 years have seen a considerable degree of rural–urban migration. This has seen, above all, the development of large Arab communities in cities in Borno – the capital Maiduguri has at least 50,000 alone[1] – though they are now found throughout cities in Nigeria.

Arabs in Nigeria are traditionally cattle nomads, part of what the anthropologist Ulrich **Braukämper1994** has called the "Baggara belt", named after the Arab

---

[1]A 1976 report by an urban planning company, the Max Lock Group (Max Lock **Group1976**) estimated that 10% of the then estimated population of 200,000 Maidugurians were Arabs. Today the population of Maiduguri is not less than one million and may be considerably larger, which proportionally would estimate an Arab population in Maiduguri alone of at least 100,000. Of course, if one included the refugee camps today, the number would be much higher.

tribe in the western Sudan (Kordofan, Darfur, see **Manfredi2010**) whose culture and dialect are very similar to those of the Nigerian Arabs. Until the very recent Bokko Haram tragedy, besides nomadism, Arabs practiced subsistence farming. As of the writing of this chapter, nearly all rural Nigerian Arabs have been forced to flee their home villages and cattle camps and are living mainly in refugee camps in NE Nigeria and neighboring countries.

Arabs first came to the Lake Chad area – whether territorial Nigeria is at this point undetermined – in the late fourteenth century. They were part of what initially was a slow migration out of Upper Egypt towards the northern Sudan beginning in the early thirteenth century, which gained momentum after the fall of the northern Nubian kingdom of Nobadia (or Maris) in the fourteenth century. All in all, Nigerian Arabic exhibits a series of significant isoglosses which link it to Upper Egypt, via Sudanese Arabic, even if it displays interesting "archaisms" linking it to regions far removed from Africa (**Owens2013**). Its immediate congeners are found in what I have termed Western Sudanic Arabic (WSA; **Owens1994a**; 1994b), stretching between NE Nigeria in the west and Kordofan in the east (**Manfredi2010**). When properties of NA are contrasted with other varieties of Arabic, it is implicitly understood that these do not necessarily include other WSA varieties. Much more empirical work is necessary in this regard, but, to give one example, many of the extended functions of the NA demonstrative described in 3.3.2 below are also found in Kordofanian Arabic (**Manfredi2014**). Moreover, where thoughout the Sudanic region as a whole any given isogloss lies is also an open question, as is the issue, the degree to which the contact-induced changes suggested here represent broad areal phenomena. As my own in many cases detailed data derives from Nigerian Arabic, I limit most observations to this area. NA itself divides into two dialect areas, a western and an eastern one that I have also termed Bagirmi Arabic since it is of a dialect with Arabs in the Bagirmi-speaking region.

In Borno Arabs are probably the largest minority ethnic group, albeit a minority. The entire area bordering Lake Chad, both to the east and to the west, is dominated by Kanuri-speaking peoples (Kanembu in Chad). This was a domination which the Arabs already met in their first migrations into the region, both a political and a linguistic domination. As will be seen, this has left dramatic influences in some domains of Nigerian Arabic, while leaving others untouched.

While until about 1970 Kanuri was the dominant co-territorial language, Arabs in the Lake Chad area have been in close contact with other languages/ethnic groups as well, for instance Fulfulde, Kotoko (just south of Lake Chad) and Bagirmi (south of Ndjammena in Chad). Furthermore, Kanuri established itself in Borno

in an area already populated by Chadic languages speakers, so it as well was probably influenced by some of the co-territorial languages Arabs met. **Since1970** Hausa has become the dominant lingua franca in all urban areas in northeastern Nigeria (indeed throughout the North). In a sample of 58 Maiduguri speakers for instance (**Owens1998**; 2000: 324), 50 professed knowing Hausa, and 46 Kanuri. In the only study of its type, Broß (**??**) shows that urban Maiduguri Nigerian Arabs have a high degree of accuracy for a number of complex variables in Hausa, while using a similar sample, in one of the few interactional studies available, **Owens2002** also documents a high multi-lingual proficiency between Arabic and Hausa, and for some speakers, English. How such micro-studies can be interpreted against the over 400 years of NA contact with area languages remains a question for the future. Rural areas have not yet experienced such a high penetration of Hausa. In a second, rural sample consisting of 48 individuals, only sixteen self-reported knowing Hausa versus forty Kanuri. Note that as of the 1990's, there were still a considerable number of monolingual Arabic speakers, particularly in the area along the Cameroonian border which among Nigerian Arabs is known as the Kala-Balge region.

While Standard Arabic (Classical Arabic) has always been a variety known among a small educated elite in Borno (of all ethnic backgrounds), along with Hausa it has gained considerable momentum in recent years. Whereas traditionally Classical Arabic, as a part of Koranic memorization, has always been a part of Arabs' linguistic repertoire, it is only since about 1990 that the teaching of Standard Arabic as a school subject has spread oral fluency in this variety.

To this point conditions have been described which, on paper at least, would favor influence via borrowing. Nigerian Arabs as a linguistic minority tend to be bilingual, and, it may be assumed, have had a history of bilingualism in Kanuri and locally other languages going back to their first migrations into the region. Equally, however, Nigerian Arabic society has itself integrated other ethnic groups creating conditions of shift to Arabic. According to Braukämper's thesis the very basis of Nigerian Arab nomadism is cattle nomadism based on a Fulani model. This is said to have arisen around the mid seventeenth century as Arabs coming from the east met Fulani moving west. Today there is very little Fulfulde spoken in Borno or Chad, so it may be surmised that the result of the Fulani-Arab contact was language shift in favor of Arabic. Furthermore, slavery was a well-established institution which incorporated speakers from other ethnic groups (see recording TV57b-Mule-Hawa in **OwensHassan2011–17**, as an instance of a slave descendant). Intermarriage is another mechanism by which L1 speakers would switch to Arabic. In contemporary Nigeria, intermarriage in fact

tends to favor Arab women marrying outside their group, rather than marriage into Arab society, though there is no cultural proscription of the practice, and such practices tend inter alia, to be influenced by the relative prestige and power of the groups involved. Today Arabs are dominated politically by the Kanuri, though there are eras, e.g. the period of Kanemi in the mid nineteenth century, or the rule of Rabeh at the beginning of the twentieth, when Arabs were more dominant and perhaps had greater access to marriage from outside groups. I will return to these summaries in section 4.

The data for this article comes from long years of working on Arabic in the Lake Chad region. More concretely, a large oral corpus of about 400,000 words (**OwensHassan2011–17**) forms the basis of much of the research, and this corpus will be referred to in a number of places in the article. When a form is said to be rare, frequent, etc., these evaluations are made relative to what can be found in the corpus. All examples come from this corpus, the source of the recording in the data bank indicated by the number in brackets at the end of the example.

## 2   2. Contact and historical linguistics

Language contact is an integral part of historical linguistics. In the case of Arabic, the history of Arabic has different interpretations, so it is relevant here to very briefly reiterate my own views (**Owens2006**). All varieties of contemporary Arabic derive from a reconstructed ancestor or ancestors. Whether singular or plural is a crucial matter, but one answered legitimately only within historical linguistic methodology (see e.g. **Retsö2013**, who appears to favour the plural). As is usually accepted (perhaps not by some working within grammaticalization theory, e.g. **HeineKuteva2011**, see fn. 2 below), historical linguistics operates at the juncture of inheritance and contact and examines change due to internal developments and change due to contact. In the case of Arabic, contact extends well into the pre-Islamic era (**Owens2013**; 2016a; to appear).

Furthermore, it operates at the level of the speech community and Arabic has and had many speech communities, each with its own linguistic history. The history of speech communities is not co-terminous with political history, usually not with the history of individual countries, or even with cultural entities such as a nomadic life style. It follows that Arabic linguistic history is quite complicated, its large population being the product of and reflecting many individual social entities.

Any individual contemporary Arabic speech community therefore lies at the end of many influences. Interpreting whether and when a particular change

occurred due to contact is anything but straightforward, as I will discuss very briefly in the following phonological issue.

Ostensibly NA shows the loss of *θ:

(1)   *θ > t, *Θawr > toor 'bull'

or in the eastern area:

(2)   *θ > s, sōr 'bull'

There is no space to go into the detailed historical linguistic arguments here, but it would be incorrect to assert that these changes, quite plausibly originally due to contact, took place in the territorial NA or WSA region. This can be seen inter alia in the fact that all of Egyptian Arabic and all of the Sudanic region including the WSA area has (**??**). Whenever the shift occurred, it was well before Arabs came to the Sudanic region, let alone Nigeria. The changes in (**??**) and, I would argue, (**??**) as well, are part of the historical linguistics of ancestral Sudanic Arabic, but the changes themselves are antecedent to Arabic in the Sudanic region and therefore are not treated here.

# 3   3. Contact induced changes

### 3.1   3.1 Phonology

Excluding cases like (**??**)–(**??**) on methodological grounds, other than marginal effects due to borrowing, discussed briefly in 3.2, there are no significant instances of contact-induced phonological change limited only to NA. Two changes confined to all or part of the WSA region can be suspected, however.

Throughout Nigeria, Cameroon and most of Chadian Arabic *ħ/ʕ have depharyngealized.

(3)   *ḥ/ʕ > h/ʔ
      *ḥilim* 'dream' > *hilim*
      *gaʕad* 'stay, sit' > *gaʔad*

As a set, the change is attested only in this region. Moreover, the area it is attested in begins by and large in the region where Arabic fades into minority status.

A second candidate for a local WSA innovation is the reflex of *ṭ, which is a voiced, emphatic implosive /ɗ/. The implosive /ɗ/ is also found in Fulfulde as well

as other possible contact languages such as Bagirmi, which, as noted above, are one source of shifters to Arabic. Manfredi (2010: 44 and personal communication) notes that /ɗ/ is an allophonic variant in Baggaara Arabic.

The status of one phoneme, /č/, is still open. It is fairly frequent (about 100 entries in the dictionary, in progress, begin with /č/). In a minority of cases an Arabic origin is certain or likely, e.g. *čāl* 'come' (eastern variant) < *tāl and perhaps *čatt* 'all', < *šattā 'various', with [š + t] > /č/ recalling some Gulf dialects *i-čūf* 'you see'. /č/ is never a reflex of *k. However, most instances of /č/ are still unaccounted for (e.g *ču* 'very red', *čāqab* 'wade through').

All in all then, there has not been a great deal of fundamental phonological change due to contact. Note that NA maintains all inherited emphatics, and probably inherited its phonemically contrastive emphatic *ṃ, *ṛ and perhaps its *ḷ as well.

## 3.2   3.2 Loanwords

Despite its long period as a minority language in the Lake Chad region, NA has only a modest number of loanwords (see **Owens2000** for a much more detailed treatment of all aspects of loanwords in the classical sense). In a token count based on about 500,000 words, only about 3% of all words were loans. On a type basis the percentage rises considerably, though still is far from overwhelming. Table **??** presents loanword provenance data from a dictionary in progress, with about 8,500 entries (excluding proper names).

Table 1: Loanwords in NA, types, N = 1263

| Language | Types |
| --- | --- |
| English | 509 |
| Hausa | 255 |
| Kanuri | 252 |
| Standard Arabic | 212 |
| French | 21 |
| Fulfulde | 12 |
| Kotoko | 2 |

The figures in Table **??** are probably a slight underestimation as there are about 60 words, like *bazingir* 'soldier of Rabeh' which clearly are not of Arabic origin but whose precise origin has not been found.

There are many interesting issues in understanding the loanwords, a few of which I mention very cursorily here. The semantic domains differ from source to source. SA, for instance, has mainly learned words. Kanuri covers a fairly wider spectrum, and strikingly includes a large number of discourse markers and conjunctions, on a token basis. *dugó* 'then, so' (K < * *dugó*) for instance has something in the range of 630 occurrences and *yo, yō, iyō* 'so, okay, aha' (**??**). In **Owens2000** discourse particles and conjunctions are shown to make up no less than 23.3% of all loanword tokens in the sample. It is noticeable that although a few Hausa discourse marker tokens (*to* 'right, okay, so') do occur, there are hardly 20 in all, this indicative of the much shorter time span Hausa has been in large-scale contact with NA as compared to Kanuri.

The question of "origin" has two aspects, one the ultimate origin, the other how it got into NA. *bel* 'belt' is ultimately of English origin, but the same word is also found in Hausa (*bel*) and in Kanuri (*bêl*). Given that both of these languages are dominant ones, it is likely that *bel* entered NA from one of these, not directly from English. The statistics above are the ultimate origin. The medial origin (travel words) is much harder to trace. Using the corpus it is possible to discern likely paths. For instance, NA *sanāʔa ~ saɲa* 'trade, occupation, profession' is cognate with both Standard Arabic *ṣinaaʕa* 'art, occupation, craft' and Hausa *sanāʔā* 'trade, craft, profession'. Considering the distribution of *saɲa* among speakers who have no knowledge of Standard Arabic, it is likely that the word reached NA via Hausa.

Non-Arabic phonology will often be maintained in the loanword. However, as can be discerned from loanwords of higher frequency, usually there is variation between retention of the source phoneme and adaptation. For instance 'police' comes in two forms, *polīs* and *folīs (**Owens2000**)*. The /p/ variant occurs in 19 tokens distributed among eight speakers, the /f/ in 18 tokens among six speakers. Inspection of the statistics shows only a tendential bias towards /f/ among women and villagers. Both variants appear therefore to be widespread. Note in this case that variation between /p/ ~ /f/ is also endemic to Kanuri, so it is likely here that the variation itself was borrowed.

## 3.3   3.3 Syntax

There are three strong candidates for contact-induced change in the syntactic domain, word order, ideophones and an expansion and realignment of demonstratives.

### 3.3.1  3.3.1 Word order and ideophones

The NA has only two pre-noun modifiers, *gōlit* 'each', *kunni* 'each'.

(4)  gōlit ʔīd      nu-lumm-u
     each holiday 1-gather-PL

     'We would gather at each festival.' (??)

Otherwise NA is head-N-initial, which means that 'all' *čatt* and 'how many' *kam* are post-N, while demonstratives only have a post-N position (as in Egyptian Arabic).

(5)  nu-mš-u be    ʔaxuwāt-na čatt-ina
     1-go-PL   with sisters-1PL   all-1PL

     'We go with our sisters, all of us.' (??)

(6)  ta-ğīb       ḍahaḅ kam
     2MSG-bring gold    how.much

     'How much gold do you bring?' (??)

The post-nominal-only demonstrative would have been inherited from Egyptian Arabic. *čatt* 'all' mirrors the post-nominal alternative for *kull*, both taking a pronoun cross-referencing the head noun. Therefore, strictly speaking the only innovation is the post-nominal position of *kam* 'how many', and an argument could be made that internal analogies lead NA towards a more consistent head-first noun-phrase order. By the same token, Kanuri is also consistently head-first order in the NP, so it could be that contact with Kanuri accelerated an inherited trend.

The numeral phrase has undergone considerable re-structuring. From twenty upwards the order is decades + ones.

(7)  talātīn haw wāhid
     thirty  and one

     'thirty one'

Though inherited teens do occur, the usual structure is ten + ones.

(8)  ʔasara haw wāhid
     ten     and one

     'eleven'

This order mirrors that of Kanuri (**Hutchison1981**), and indeed that of most languages in the immediate Lake Chad area. Uzbekistan Arabic has the same numeral order and structure as NA, and in these cases independent contact events are likely the reason for the shift from an inherited structure.

A new syntactic category (for Arabic), that of ideophones, is described in detail in **OwensHassan2004** (see *tul* in (??) below). To date in the dictionary of NA (in progress) there are 342 ideophones, about 4% of the lemma total.

### 3.3.2  3.3.2 Demonstratives

Formally NA demonstratives reproduce their inherited forms, and therefore are virtually identical to paradigms found in various Egyptian dialects, except that in consonance with NA morphology, feminine plural has a distinct form, which most Egyptian dialects have neutralized (see Table ??).

Table 2: NA demonstratives

|  | Near | | Far | |
|---|---|---|---|---|
|  | Singular | Plural | Singular | Plural |
| Masculine | da | dōl | ḍāk | ḍōlak |
| Feminine | di | dēl | ḍīk | ḍēlak |

As with all Arabic demonstratives, NA demonstratives are used both as modifiers and pronominally. The traditional, inherited functions are entity referential (*al-bēt da* 'this house'), and propositional anaphoric (*ʔašān da* 'because of this', where 'this' references an introduced proposition).

Additionally, however, the demonstratives occur in a number of contexts which either are not attested at all, or are attested only on an extremely infrequent basis in other Arabic dialects. I summarize these here.

3.3.2.1  1. Marking the end of dependent clauses, whether relative, conditional or adverbial. Usually *da* is the default form in this function, though in the case of relative clauses the demonstratives often agrees with the head noun.

(9)   Conditional clause
[kan gul     ba-lkallam      kalām-hum     da]     ma bukūn
[if     said.1SG 1SG.IND-speak language-3MPL DEM.M] NEG possible
'If I said I speak their language, it is not possible.' (??)

(10) Relative clause
ba-lkallam le-əm    be    l-luqqa l-bi-yarif-ū-ha
1SG-speak   to-3MPL with DEF     language
di
REL-3M.IND-know-PL-3FSG DEM.F
'I speak to them in the language which they know.' (??)

3.3.2.2 2. Text referential, cataphoric. *da* is used cataphorically to foreshadow a propositional expansion. In (??) the speaker is asked how he farms. Instead of answering directly, he introduces his answer with the cataphoric use of *da*, which is then expanded upon in the following independent proposition.

(11) kēf   ti-hērit
how 2MSG-farm
'How do you farm?'

(12) ba-harit       da,    al-hirāta    l-wād-e tul di    d-duxun
1SG.IND-farm DEM.M DEF-farming DEF-one-f only DEM.F DEF-millet
'How I farm? The one type of farming is only millet.'

3.3.2.3 3. Deictics.

A number of deictic words, mainly adverbs, are marked by demonstratives, in this case nearly always *da*. The deictics include *hassa* 'now', *dugut* 'now', *wakit/waqit* 'now', *tawwa* 'previously, formerly', *hine/hinēn* 'here', *awwal* 'first, before', *gabul* 'previously, before, *baʔad* 'afterwards', *alōm/alyōm* 'today', *amis* 'yesterday', *al-bāre* 'yesterday evening', *ambākir* 'tomorrow', *bukura* 'day after tomorrow', *məṇṇaṣabá* 'in the morning', *qādi* 'there', *hināk* 'there', *hajira* '(a place) away from here', *bil-hēn* 'much'

(13) hajira da      ma  mašē-t
away  DEM.M NEG go-1SG
'I didn't go away anywhere.' (45 Mag)

(14) albāre     da    as-sarārīk daxal-o
yesterday DEM.M DEF-thieves enter-3MPL
'Yesterday evening thieves broke in.' (??)

3.3.2.4  4. Demonstratives mark pronouns, in this case often agreeing with the pronoun in terms of number and gender, and other demonstratives, where usually *da* occurs.

(15)  inti   di      ǧīb-i    le-i    š-suqúl    da
      you.f DEM.F bring-f to-1SG DEF-thing DEM.M
      'You there bring me the watchamacallit.' (??)

(16)  ʔard gaidam dōla     da       kula ʔarab
      land Geidam DEM.MPL DEM.MSG also  Arab
      'In the land around Geidam and the like are also Arabs.' (??)

Basic attributes of these expanded functions can be given in cursory manner.

Frequency. The occurrence of demonstratives in these functions on a token basis is high. For instance, there are 887 tokens of *qādi* 'there' in the corpus, of which 108 or 12% are marked by *da*. The highest percentages of demonstratives in these functions occur with the dependent clauses and the 3SG pronouns *hu* 'he' and *hi* 'she'.

For *hu*, nearly 25% of all tokens occur with *da* (586/2407 24.3%).

As far as the four innovative functions summarized above, a sample of 1318 tokens of *da* gathered from an arbitrary selection of 45 texts in the corpus reveals the data presented in Table ??.

Table 3: Functions of da in NA

| Function | Percentage of total |
|---|---|
| Inherited functions | 53.4 |
| Entity referential: 557 | 42.3 |
| Proposition-anaphoric: 146 | 11.1 |
| Innovative functions | 46.7 |
| Cataphoric-propositional: 95 | 7.2 |
| Dependent clause: 246 | 18.7 |
| Adverbs/deictic: 158 | 12 |
| Pronouns, demonstratives: 116 | 8.8 |

While the inherited referential functions constitute the largest single class, they make up only 53% of the total. The remaining 47% are functionally innovative.

The syntactic, pragmatic and semantic nuances of using or not using the demonstratives in these innovative contexts have yet to be worked out. Two examples illustrate different ways the innovative functions are integrated with other elements of the grammar.

Syntactically, for instance, based on the sample described above, *da* marks the end of about 30% of all conditional clauses. When it does not occur, its final clause boundary marking position commutes with an alternative pragmatically-marked element, such as the discourse marker *kula* 'even'.

(17)  kan   qayyar-t-a          kula
      if    change-2MSG-3MSG DM

      'even if you changed it' (??)

(No tokens of \**kula da* closing a conditional clause occur in the corpus).

Pragmatically there are many instances where *da* has a focusing function, as in the following, where a mixed linguistic region 'here' is contrasted with another 'there', which is linguistically homogeneous.

(18)  nās   gadé   gadé   kula hinēn katīrīn fi   amma [qādi **da**]
      people different different DM   here  many  exist [there DEM] type-1PL

      nafar-na nafara wāhid
      type     one

      'Here there are a lot of different (types) but [over there] there is just our one ethnic group.'

The functions outlined in Table ?? are therefore both of high frequency and are systematically embedded in the syntax and pragmatics.

It should be intuitively clear that the functions in examples (??)–(??) are innovative in their systematicity relative to other varieties of Arabic. To show this in detail it would, however, be necessary to look at large-scale corpora of other Arabic dialects. This can very briefly be done with Egyptian Arabic, which, as noted above, is an ancestral homeland of NA. The EA corpus is from ***LDC Callhome*** **(??)**, **Nakano1982**, **BehnstedtWoidich1987**, and **WoidichDrop2007**, comprising about 417,000 words. It is thus of comparable size to the NA corpus. In this corpus there do occasionally occur collocations of pronoun + demonstrative in the same contexts as illustrated in (??), in particular as in (??).

(19)  hiwwa da      lli   mawgūd ʕand-ina
      it.M    DEM.M REL present  at-1PL

      'That is what we have.'

It clearly, however, has a different functionality from NA pronoun + demonstrative. In Egyptian Arabic (EA) the construction consistently is anaphoric to a previous proposition or situation, as in (??), where it introduces a previously-established topic to a following descriptive qualification. In 11 of the 58 tokens in the EA corpus it is followed by a relative clause, as in (??). Most tellingly, there are 2,677 *huwwa* (~ *hu, hū, hiwwa, hūwa*) tokens, of which only 58, or 2% are followed by *da* (~ *dah, dih, deh, dī*). This compares to the nearly 25% *hu + da* tokens in NA noted above. Moreover, in the NA sample, no tokens of *hu da* are followed by a relative clause. In this same statistical vein, the total number of singular proximal demonstratives in NA amounts to 16,774 tokens (14,591 *da*, 2,183 *di*). In the EA corpus there are "only" 8,239 (4,996 *da*, 3,243 *di*). Given that the corpora sizes are comparable (EA in fact a little larger), the demonstratives in NA are vastly over-proportional. This preponderance is due to *da*. Clearly there is a case to be answered for: what accounts for the vastly higher frequency of the 3MSG demonstrative in NA? Recall, in answering this question, that behind the simple statistical comparison is a fundamental historical one as well. Ancestral NA came from ancestral EA. The initial populations, it needs to be assumed, had a demonstrative system like that of EA, and the majority of NA demonstrative tokens (see Table ??) still reflect this system. A blunt historical linguistic question is what caused the vast shift in frequencies.

From these initial, basic observations, it does not appear that the greatly expanded functionality of the demonstrative in NA can be explained by an increasing grammaticalization of the demonstrative.[2] This follows from two observations. First the expanded functions of the demonstrative in Table ?? are, with the exception of the boundary-marking of dependent clauses (??), not those associated with the grammaticalization of demonstratives (e.g. the 17 trajectories of demonstratives in **Diessel1999**). Secondly, NA and EA split over 400 years ago. One of the branches, represented by NA, underwent the considerable changes outlined here, whereas the other branch, EA, probably did not change at all (i.e. sentences such as (??) were probably present in EA in 1200, and before).[3] There is thus no natural or inherent tendency for demonstratives to expand as in Ta-

---

[2]I do not at all agree with **HeineKuteva2011** that changes due to contact can be assimilated to a type of grammaticalization process, so the following contact-based account is independent of grammaticalization. Grammaticalization, in the original Meilletian sense, pertained only to internally-motivated changes.

[3]Cf. Damascus, which has an identical construction as the EA, *hāda huwwe llimawžūd ʕindna (mā fī šī tāni* 'there's nothing else). There are parallels in Classical Arabic as well, so this type of construction is probably proto-Arabic. If so, it only heightens the degree to which NA has innovated away from an original, stable structure.

ble **??**. It can thus be safely assumed that the expanded functionality of the NA demonstrative was due to contact.

In fact, there is a large amount of prima facie evidence supporting this supposition. However, as is so frequently the case when one suspects pattern (metatypical)-type contact influence which is probably centuries old, support for the position will be indexical. Moreover, in the current case one is most probably dealing with a large-scale areal phenomenon in the Lake Chad area (and perhaps beyond) which encompasses well over 100 languages. In this summary article it will therefore have to suffice to rather peremptorily indicate that throughout the region there is a referential marker, sometimes a demonstrative, sometimes an article-like element, sometimes an element with both demonstrative and article-like properties which consistently has the distribution of (**??**)–(**??**). Some languages have a better fit than others, and of course, they will differ in detail in their language-internal functionality. A basic pattern is illustrated in (**??**) with Kanuri (**Hutchison1981**: 47, 207, 218, 234, 241, 270), and summary references are made for Bagirmi, Wandala and Fali. So far as is known, Fali and Wandala had no significant contact with NA or its WSA relatives.

The Kanuri determinative -*də* has the following functions, its basic being (**??**).

(20)  Kanuri (Nilo-Saharan/Saharan)
      anaphoric entity reference;

      a.  obligatorily ends RC and optionally many adverbial clauses; = (**??**), (**??**)
      b.  pronoun focus; = (**??**)
      c.  marks adverbs; = (**??**), (**??**)

The only Kanuri structure "missing" from the list appears to be the propositional cataphoricity illustrated in (**??**).

Wandala (**Frajzyngier2012**: 507–34, 603). Wandala has two morphemes: -*na* which is broadly glossed as a determiner and -*w* 'that'. -*na*, besides marking entity reference, obligatorily marks the ends of a relative clause, and optionally a conditional (=(**??**), (**??**)), it occurs as obligatory element in certain time/place adverbs (=(**??**), (**??**)); it is part of the previous mention marker *ŋán-na; ŋán* itself is said to originally be a third person singular pronoun, so there is a structural parallel to *hu* + *da*. -*w* functions as a topic marker that marks pronouns (=(**??**)).

In Fali (Niger-Congo/Adamawa) the demonstratives *gi/go* also obligatorily mark the end of relative and conditional clauses (=(**??**), (**??**)), subject focus (=(**??**)) and occur with some adverbs(=(**??**), (**??**)).

In Bagirmi a "determiner particle" *-na* is a constitutive part of the demonstrative *enna* < *et-na* 'this', and *-na* alone obligatorily marks the end of relative clauses, and can emphasize pronouns, adverbs and entire sentences (**Stevenson1969**: 40, 51, 54).

Areal features typically are not sensitive to language family, and this appears to be the case in this brief exemplification. Kanuri and Bagirmi are Nilo-Saharan, Wandala Chadic, Fali Niger-Congo and Arabic is Semitic. Only Wandala and Arabic are very distantly related genetically. Nonetheless, in all of the languages there is a deictic–referential marker (demonstrative, determiner, demonstrative–determiner) which, besides a "classic" deictic or anaphoric function, surfaces in an extended range of identical (cf. marking boundary of dependent clause) or similar (pronouns, adverbs)[4] functions. These extended functions are precisely those which distinguish NA from other varieties of Arabic. The case for contact follows from two directions: in certain (not all) respects NA deviates markedly from a putative ancestral source shared with EA, and where it does its deviation corresponds broadly to analogous categories in co-territorial languages.

## 3.4   3.4 Semantics

The innovative distribution of the NA demonstratives is striking for the degree to which it appears to have raised the overall demonstrative token count, relative to EA. Discerning its presence in a text, however, is a straightforward matter. A much subtler, but no less pervasive instance of contact-based change pertains to idiomaticity. Like the demonstrative, this has a semantic and a formal aspect. Semantically meanings emerge which are, for Arabic, unique, as in the following.

(21)  rās    al-bēt
      head DEF-house
      'roof'

(22)  nādim  rās-a
      person head-3MSG
      'an independent person, person of his own means'

---

[4]The comparativist is limited to the extant reference grammars. These are in many instances excellent. Still, I suspect that they understate the flexibility of distribution of elements such as the deictic marker discussed here. Mea culpa, in Owens (1993: 88, 221, 235) the extended functions of the demonstrative described in this article for NA were treated in disparate sections, with no overall focus.

(23)  tallaf-o      gaḷb-i
      spoil-3MPL heart-1SG
      'They angered me.'

(24)  gaḷb-a      helu
      heart-3MSG sweet
      'He is happy.'

Formally the idioms are distinctive (as Arabic collocations) in bringing together lexemes which in other dialects would hardly co-occur, like [tallaf + gaḷb] or [gaḷb + helu]. The idiomatic meanings of the keywords (e.g. *tallaf, gaḷb*) are, in usage terms, often the typical usage for a given lexeme. In the NA corpus, for instance, of 101 tokens of *gaḷb* 'heart' all of them, 100%, are idiomatic. There is no reference to a physical heart. Similarly, *rās* is 80% idiomatic (247/308 tokens; Ritt-**BenmimounEtAl2017**: 53). Thus, while idiomaticity has been consistently ignored as a theoretical issue in historical linguistics in general and in Arabic in particular, on a usage basis it is an integral aspect of understanding the lexical texture of the language.

Here as well NA is strikingly different from EA, as again can be determined from corpora-based comparison. In general, though both NA and EA share idiomatic keywords, for instance in both *gaḷb* and *ṛaas* are frequent, their meanings and their collocational environments hardly overlap. For instance, in the EA corpus there are 110 tokens of *gaḷb/ʔalb* 'heart', of which 102 or 93% are idiomatic. This percentage closely parallels that of NA idiomatic *gaḷb*. The typical EA collocate of idiomatic *ʔalb*, however, is very different. The most frequent meaning is 'center of X', *ʔalb il-baḥr* 'middle of the sea'. This meaning is entirely lacking in NA, and consequently collocates like !*gaḷb al-bahar* (! = collocationally/semantically odd) are also lacking.

How different NA idiomaticity (meaning and collocational environment) is from EA was shown recently in Ritt-**BenmimounEtAl2017**. There a three-way comparison was conducted between EA, southern Tunisian Arabic and NA looking at three idiomatic keywords frequent in all three dialects, *ṛās, gaḷb* 'heart', and *ʕēn* 'eye'. EA and southern Tunisian, though separated by a longer period of time (ca. 1035–present) than EA–NA (ca. 1300–present), showed a much higher identity of idiomatic structure than EA–NA (or NA–southern Tunisian). Both EA (**??**)a and Tunisian (**??**)b, for instance maintain the same lexemes, same structure, same idiomaticity in a highly specific meaning.

(25) ḥaṭṭ        ṛās-u         fi t-turāb
     put.PRF.3MSG head-3MSG in DEF-ground
     'He humiliated him.'

(26) ḥaṭṭ-l-a                  ṛās-a       fi t-ṭṛāb
     put.PRF.3MSG-DAT-3MSG head-3MSG in DEF-ground
     'He humiliated him.'

These are nonsensical, or literal collocations in NA.

The comparison between EA and southern Tunisian Arabic serves as a similar baseline to comparing the overall demonstrative frequencies between EA and NA. The same question occurs. Why is NA different?

In this case the answer is even clearer than with the demonstrative. Essentially, NA has calqued its idiomatic structure (meaning and collocation) from Kanuri. The Kanuri of (??)a and (??)b for instance are as in (??).

(27) kəla  fato-be
     head house-of
     'roof'

(28) kam    kəla-nzə-ye
     person head-3MSG-of
     'an independent person, person of his own means'

A 'roof' in both languages is the 'head of a house', an independent person is a 'person of his head', and so on, for something in the range of 70–80% of all the approximately 340 idioms studied (see **Owens1996**; 2014; 2015; 2016b for details).

In summary, a large part of NA lexical structure is, as it were, not Arabic, but rather, as termed in **Owens1998**, part of the Lake Chad idiomatic area. This identity, however, exists only at a semantic and collocational level. In their basic meaning, and their phonology, morphology and syntax even in the context of idioms (**OwensDodsworth2017**), the constituent lexemes *rās*, *bēt*, *tallaf*, *gaḷb* etc. in NA are indistinguishable from any variety of Arabic at all.

There doubtless remains a good deal more systematic, contact-based correspondence between NA and languages of the Lake Chad area to be explored. The influence on NA is significant.

# 4  4. Conclusion

According to the historico-demographic background to Nigerian Arabic, Nigerian Arabic did and does live with co-territorial languages, particularly Kanuri, today increasingly with Hausa, and in the past, Fulfulde and other smaller languages. NA bilingualism should, presumably, manifest itself in borrowing. Equally, NA speech communities have incorporated speakers of other languages into its fabric. The expectation here is that NA would be influenced via shift (imposition) from other languages.

In the domains summarized here, it is hard to discern a clear correlation between linguistic outcome and type of contact. There has been some phonological change, which in Van Coetsem's (1988; 2000) model is suggestive of change via shift (imposition), but the influence is limited to the features discussed in §3.1. What I believe is more striking than the contact-induced phonological change is the maintenance of inherited structures. NA still maintains a robust series of emphatics, has a non-reductive syllable structure reminiscent of, inter alia Tihāma varieties, has "classic" distinguishing syllable structure attributes such as the *gahawa* syndrome (*ahamar* 'red') and the *bukura* syndrome (*bi-ǧiri* 'he runs'), to mention but a few. If the changes in (??)–(??) are due to imposition, it is equally clear that the "imposers" otherwise learned/learn a very normal Arabic.

Classic borrowing is moderate. The fact that discourse markers and conjunctions are token-wise frequent suggests that speakers were/are conversant in both Kanuri and Arabic. This does not, however, indicate whether these borrowings arose through imposers or borrowers. Moreover, to complicate matters even more, assuming Kanuri to have the widespread lingua franca in the past, it would not need to have been native Kanuri speakers who "imposed" the Kanuri into Arabic. Speakers of Fulfulde, Kotoko, Malgwa or other languages would have been involved as well. As shown in **OwensHassan2010**, discourse markers are prevalent in codeswitching, which here would be conducted by Arabs codeswitching between Arabic and Kanuri. From this scenario the discourse markers entered as borrowed elements.

The interpretation of demonstratives and idiomatic structure is equally ambiguous. The easiest development to envisage is L2 Arabic speakers imposing their L1 Kanuri, Fulfulde etc. usage onto their L2 Arabic. What makes this interpretation attractive is that it explains why in both cases such a massive importation of non-Arabic structure came into Arabic. As the name implies, these speakers could simply have imposed their own semantics and collocational alignment onto Arabic. Equally, however, it is not impossible that L1 Arabic speakers, fully

bilingual in Kanuri and/or other languages simply shifted their Arabic usage to accommodate to their L2. Full fluency implies knowing idiomatic structure and the use of demonstratives, which the Arab borrowers could eventually incorporate into their own Arabic.

The only obvious common denominator to these musings is that the speakers would have been highly fluent in their respective L2s, whether L2 Arabic speakers shifting to Arabic or L1 Arabic speakers fluent in Kanuri or other languages borrowing from their L2. The issue is only partly who the L1 and L2 speakers are. It is equally how well the populations knew/know Arabic/other languages, and how the high level of fluency produces the results shown.

Adding to the interpretive problem is that neither of the domains, idiomaticity or the expansion of demonstratives as it occurred in NA, have a comparative basis. Idiomaticity in the recent western linguistic tradition has been all but entirely subordinated to metaphor theory (**LakoffJohnson1999**; see **Haser2005** for one critical perspective), it has received very little principled historical interpretation, and what work has been done (e.g. **Sweetser1990**) tends to follow a Lakoffian paradigm and to be confined to European languages and to societies quite different from that of Nigerian Arabs. As far as demonstratives go, the little work that has been done on the languages co-territorial with NA (e.g. **Kramer2014**: 141 on Fali), assume a grammaticalization of demonstrative usage *ab novo* via grammaticalization processes. Assuming such a perspective for the development of NA gives the lie to this simple assumption for the following reason. It would need to explain why the grammaticalization process did not take place in EA or other Arabic varieties, but did in NA, which is spoken in an area where the co-territorial languages, historically antecedent to Arabic, have the structures which NA acquired. If change via contact is the only plausible explanation for NA, it equally needs to be entertained for any language in the Lake Chad region.

Given so many open variables, it might be interesting to approach the issue from the opposite perspective, namely, what parts of language were not influenced by contact. Most of phonology was not, morphology hardly at all, syntax to a degree, basic vocabulary little.[5] This minimally implies that if the contact changes were due to shift, the shifters in other domains (those where they did not impose idiomaticity or demonstrative usage) acquired a native-like competence in Arabic. In this respect it might be easier to envisage L1 Arab borrowers maintaining these structures, and borrowing idiomaticity/demonstrative usage via their L2.

---

[5]A Swadesh 100-word list gives something in the range of 79–83% cognacy with other varieties of Arabic.

At the end of the day I think the range of questions evoked far surpasses the ability of currently-formulated linguistic theories of contact or language change, whether based on sociolinguistic or on cognitive perspectives (**Lucas2015**) to provide profound insight into how the obvious, and in some cases pervasive influence on NA via contact came about. It would be more fruitful to turn the question around and ask how rich data bases such as exist for NA, EA and some other Arabic dialects inform the overall issue of change via contact.

```
%%move bib entries to  localbibliography.bib


@book{Behnstedt1987,
address = {Wiesbaden},
author = {Behnstedt, Peter  and  Manfred Woidich.},
publisher = {Reichert},
title = {\textit{Die} \textit{Ägyptisch-arabischen} \textit{{Dialekte},} \textit
year = {1987}
}




@article{Braukämper1994,
author = {Braukämper, Ulrich},
journal = {\textit{Sprache} \textit{und} \textit{Geschichte} \textit{in} \textit
pages = {13–46},
title = {Notes on the origin of Baggara Arab culture with special reference to t
volume = {14},
year = {1994}
}




@incollection{Broß2007,
address = {Hamburg},
author = {Broß, Michael},
booktitle = {\textit{Languages} \textit{in} \textit{{African}} \textit{urban} \t
editor = {Gudrun Miehe, Jonathan Owens \& Manfred von Roncador},
pages = {51–74},
publisher = {LIT},
```

title = {L2 speakers, koineization and the spread of language norms: {H}ausa in {Ma
year = {2007}
}



@book{Coetsem1988,
address = {Dordrecht},
author = {Coetsem, Frans van.},
publisher = {Foris},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two} \t
year = {1988}
}



@book{Coetsem2000,
address = {Heidelberg},
author = {Coetsem, Frans van.},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theory}
year = {2000}
}



@article{Diessel1999,
author = {Diessel, Holger},
journal = {\textit{Linguistic} \textit{Typology}},
pages = {1–46},
title = {The morphosyntax of demonstratives in synchrony and diachrony},
volume = {3},
year = {1999}
}



@book{Frajzyngier2012,

```
address = {Berlin},
author = {Frajzyngier, Zygmunt.},
publisher = {Mouton de Gruyter},
title = {\textit{{Wandala}}},
year = {2012}
}




@book{Haser2005,
address = {Berlin},
author = {Haser, Verena.},
publisher = {Mouton de Gruyter},
title = {\textit{Metaphor,} \textit{metonymy} \textit{and} \textit{experientiali}
year = {2005}
}




Heine, Bernd \& Tania Kuteva. 2011. The areal dimension of grammaticalization. I




@book{Hutchison1981,
address = {Madison},
author = {Hutchison, John.},
publisher = {African Studies Program},
title = {\textit{The} \textit{{Kanuri}} \textit{language}},
year = {1981}
}




@book{Kramer2014,
address = {Köln},
author = {Kramer, Raija.},
publisher = {Rüdiger Köppe},
title = {\textit{Die} \textit{{Sprache}} \textit{der} \textit{Fali} \textit{in}
```

```
year = {2014}
}




@book{Lakoff1999,
address = {New York},
author = {Lakoff, George  and  Mark Johnson.},
publisher = {Basic Books},
title = {\textit{Philosophy} \textit{in} \textit{the} \textit{flesh}},
year = {1999}
}




@incollection{Lucas2015,
address = {London},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{{Routledge}} \textit{handbook} \textit{of} \text:
editor = {Claire Bowern \& Bethwyn Evans},
pages = {519–536},
publisher = {Routledge},
title = {Contact-induced language change},
year = {2015}
}




@book{Manfredi2010,
address = {Naples},
author = {Manfredi, Stefano.},
publisher = {Università degli studi di Napoli L'Orientale doctoral dissertation},
title = {A grammatical description of Kordofanian Baggara {Arabic}},
year = {2010}
}
```

```
@article{Manfredi2014,
author = {Manfredi, Stefano},
journal = {\textit{Folia} \textit{Orientalia}},
pages = {27–50},
title = {Demonstratives in a {Bedouin} Arabic dialect of western Sudan},
volume = {51},
year = {2014}
}
```

```
@book{Max1976,
address = {London},
author = {Max Lock Group.},
publisher = {Westminster},
title = {\textit{{Max}} \textit{Lock} \textit{Group} \textit{report:} \textit{Ma
year = {1976}
}
```

```
@book{Nakano1982,
address = {Tokyo},
author = {Nakano, Aki'o.},
publisher = {Institute for the Study of Languages and Cultures of Asia and Afric
title = {\textit{Folktales} \textit{of} \textit{Lower} \textit{{Egypt}:} \textit
year = {1982}
}
```

Owens, Jonathan. to appear. Arabic. In Salikoko S. Mufwene \& Anna María Escobar

```
@book{Owens1993,
address = {Wiesbaden},
author = {Owens, Jonathan.},
```

```
publisher = {Harrassowitz},
title = {\textit{A} \textit{Grammar} \textit{of} \textit{{Nigerian}} \textit{Arabi
year = {1993}
}




@misc{Owens1994,
author = {Owens, Jonathan (ed.)},
number = {14},
series = {SUGIA)},
title = {\textit{Arabs} \textit{and} \textit{{Arabic}} \textit{in} \textit{the} \te
year = {1994a}
}




@article{Owens1994,
author = {Owens, Jonathan},
journal = {\textit{Sprache} \textit{und} \textit{Geschichte} \textit{in} \textit{A
pages = {85–176},
title = {{Nigerian} Arabic in comparative perspective},
volume = {14},
year = {1994b}
}




@article{Owens1996,
author = {Owens, Jonathan},
journal = {\textit{Diachronica}},
number = {2},
pages = {283–318},
title = {Idiomatic structure and the theory of genetic relationship},
volume = {13},
year = {1996}
}
```

*Jonathan Owens*

```
@book{Owens1998,
address = {Amsterdam},
author = {Owens, Jonathan.},
publisher = {John Benjamins},
title = {\textit{Neighborhood} \textit{and} \textit{ancestry:} \textit{Variation
year = {1998}
}
```

```
@incollection{Owens2000,
address = {Berlin},
author = {Owens, Jonathan},
booktitle = {\textit{{Arabic}} \textit{as} \textit{a} \textit{minority} \textit{
editor = {Jonathan Owens},
pages = {259–346},
publisher = {De Gruyter},
title = {Loanwords in {Nigerian} Arabic: {A} quantitative approach},
year = {2000}
}
```

```
@article{Owens2002,
author = {Owens, Jonathan},
journal = {\textit{Language} \textit{Variation} \textit{and} \textit{Change}},
number = {2},
pages = {173–209},
title = {Processing the world piece by piece: {I}conicity, lexical insertion and
volume = {14},
year = {2002}
}
```

```
@book{Owens2006,
```

```
address = {Oxford},
author = {Owens, Jonathan.},
publisher = {Oxford University Press},
title = {\textit{A} \textit{linguistic} \textit{history} \textit{of} \textit{{Arab
year = {2006}
}




@article{Owens2013,
author = {Owens, Jonathan},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{American} \textit{Orie
number = {2},
pages = {217–248},
title = {The historical linguistics of the intrusive \textit{{}-n} in {Arabic} and
volume = {133},
year = {2013}
}




@article{Owens2014,
author = {Owens, Jonathan},
journal = {\textit{Linguistics}},
number = {1},
pages = {125–165},
title = {Many heads are better than one: {T}he spread of motivated opacity via con
volume = {52},
year = {2014}
}




@article{Owens2015,
author = {Owens, Jonathan},
journal = {\textit{Anthropological} \textit{Linguistics}},
number = {1},
pages = {46–98},
```

```
title = {Idioms, polysemy, context: {A} model based on {Nigerian} Arabic},
volume = {57},
year = {2015}
}



@incollection{Owens2016,
address = {Leiden},
author = {Owens, Jonathan},
booktitle = {\textit{Approaches} \textit{to} \textit{the} \textit{history} \text
editor = {Manuel Sartori, Manuela Giolfo \& Philippe Cassuto},
pages = {77–101},
publisher = {Brill},
title = {Dia-planar diffusion: {R}econstructing early Aramaic–{Arabic} language
year = {2016a}
}



@article{Owens2016,
author = {Owens, Jonathan},
journal = {\textit{Language} \textit{Sciences}}},
pages = {49–69},
title = {The lexical nature of idioms},
volume = {57},
year = {2016b}
}



@article{Owens2017,
author = {Owens, Jonathan  and  Robin Dodsworth},
journal = {\textit{Linguistics}}},
number = {3},
pages = {641–682},
title = {Semantic mapping: {W}hat happens to idioms in discourse},
volume = {55},
```

```
year = {2017}
}




@incollection{Owens2004,
address = {Leiden},
author = {Owens, Jonathan  and  Jidda Hassan},
booktitle = {\textit{A} \textit{{Festschrift}} \textit{for} \textit{Manfred} \text
editor = {Martine Haak, Rudolf de Jong \& Kees Versteegh},
pages = {207–220},
publisher = {Brill},
title = {Remarks on ideophones in {Nigerian} Arabic},
year = {2004}
}




@incollection{Owens2010,
address = {London},
author = {Owens, Jonathan  and  Jidda Hassan},
booktitle = {\textit{Information} \textit{structure} \textit{in} \textit{spoken} \
editor = {Jonathan Owens \& Alaa Elgibali},
pages = {207–242},
publisher = {Routledge},
title = {Conversation markers in {Arabic}-Hausa codeswitching: {S}aliency and langu
year = {2010}
}




Retsö, Jan. 2013. What is Arabic? In Jonathan Owens (ed.), \textit{The} \textit{Ox
450. Oxford: Oxford University Press.




@incollection{Ritt-Benmimoun2017,
address = {Zaragoza},
```

author = {Ritt-Benmimoun, Veronika, Grigore Smaranda, Jocelyne Owens  and  Jonat
booktitle = {\textit{Tunisian} \textit{and} \textit{Libyan} \textit{{Arabic}} \t
} \textit{recent} \textit{developments} \textit{--} \textit{diachronic} \textit{As
editor = {Veronika Ritt-Benmimoun},
pages = {43--84},
publisher = {University of Zaragoza Press},
title = {Three idioms, three dialects, one history},
year = {2017}
}


@book{Stevenson1969,
address = {Khartoum},
author = {Stevenson, R. C.},
publisher = {Sudan Research Unit, University of Khartoum},
title = {\textit{Bagirmi} \textit{grammar}},
year = {1969}
}


@book{Sweetser1990,
address = {Cambridge},
author = {Sweetser, Eve.},
publisher = {Cambridge University Press},
title = {\textit{From} \textit{etymology} \textit{to} \textit{pragmatics:} \text
year = {1990}
}


@book{Woidich2007,
address = {Wiesbaden},
author = {Woidich, Manfred  and  Hanke Drop.},
publisher = {Harrassowitz Verlag},
title = {\textit{ilBaħariyya} \textit{--} \textit{{Grammatik}} \textit{und} \text
year = {2007}

```
}


\subsubsection{Online resources}

@misc{Canavan1997,
author = {Canavan, Alexandra, George Zipperlen  and  David Graff},
title = {\textit{CALLHOME} \textit{{Egyptian}} \textit{Arabic} \textit{Speech} \tex
year = {1997}
}

Owens, J. and Jidda Hassan. 2011-17. \textit{In} \textit{their} \textit{own} \text
```

# Abbreviations

# Acknowledgements

Chapter 5

# Contact-induced change in Jerusalem Domari

Yaron Matras

University of Manchester

Jerusalem Domari is the only variety of Domari for which there is comprehensive documentation. The language shows massive influence of Arabic in different areas of structure – quite possibly the most extensive structural impact of Arabic on any other language documented to date. Arabic influence on Jerusalem Domari raises theoretical questions around key concepts of contact-induced change as well as the relations between systems of grammar and the components of multilingual repertoires; these are dealt with briefly in the chapter, along with the notions of fusion, compartmentalisation of paradigms, and bilingual suppletion.

## 1 Historical development and current state

Domari is a dispersed, non-territorial minority language of Indo-Aryan origin that is spoken by traditionally itinerant (peripatetic) populations throughout the Middle East. Fragmented attestations of the language place it as far north as Azerbaijan and as far south as Sudan. The self-appellation *dōm* is cognate with those of the *řom* (Roma or Romanies) of Europe and the *lom* of the Caucasus and eastern Anatolia. All three populations show linguistic resources of Indo-Aryan origin (which in the case of the Lom are limited to vocabulary) as well as traditions of a mobile service-economy and are therefore all believed to have descended from itinerant service castes in India known as *ḍom*. Some Domari speaking populations are reported to use additional names including *qurbāti* (Syria and Lebanon), *mıtrıp* or *karači* (Turkey and northern Iraq) and *bahlawān* (Sudan), while the surrounding Arabic-speaking populations usually refer to them as *nawar*, *ɣaǧar* or *miṭribiyya*. The language retains basic vocabulary of Indo-Aryan origin, and shows elements of lexical phonology that place its early development within the

Central Indo-Aryan group of languages. It retains conservative derivational as well as present tense inflectional verb morphology that goes back to late Middle Indo-Aryan, alongside innovations in nominal and past tense verb inflection that suggest that the language was contiguous with the Northwestern frontier languages (Dardic) during the transition to early modern Indo-Aryan (cf. **Matras2012**).

The first attestation of Palestinian Domari is a list of words and phrases collected by Ulrich Jasper Seetzen in 1806 in the West Bank and published by **Kruse1854**. It was followed by **Macalister1914**'s (**Macalister1914**) grammatical sketch, texts and lexicon collected in Jerusalem in a community which at the time was still nomadic, moving between the principal West Bank cities of Nablus, Jerusalem and Hebron. This community settled in Jerusalem in the early 1920s, the men taking up wage employment with the British-run municipal services. In the 1940s they abandoned their makeshift tent encampment and moved into rented accommodation within the Old City walls, where the community still resides today. **Between1996** and 2000 I carried out fieldwork among speakers in Jerusalem and published a series of works on the language, including two descriptive outlines (**Matras1999**; 2011), annotated stories (**Matras2000**), an overview of contact influences (**Matras2007**), and a descriptive monograph (**Matras2012**).

A number of sources going back to **Pott1846**, **Newbold1856**, **Paspati1870**, **Patkanoff1907**, and **Black1913** provide language samples collected among the Dom of Lebanon, Syria, Iraq and the Caucasus. These are supplemented by a few more samples collected by ethnographers (cf. **Matras2012**: 15ff.) and subsequently by data collected in Syria and Lebanon by **Herin2012**. That documentation allowed me to identify a number of differences that appeared to separate a Northern group of Domari dialects from a Southern group, which latter includes the data recorded in Palestine as well as a sample from Jordan (see **Matras2012**: 15ff.). That tentative classification has since been embraced by **Herin2014**, who goes a step further and speculates about an early split between two branches of the language. To date, however, published attestation of Northern varieties remains extremely fragmented, notwithstanding recent work by **Herin2016**, while the only comprehensive overview of a Southern variety remains that from Jerusalem.

Outside of Jerusalem and its outskirts there are known communities of Palestinian Doms in some of the refugee camps on the West Bank and Gaza, as well as in Amman, where a few families sought refuge in 1967. Numbers of speakers were very low in all these communities already in the mid 1990s and the language was only in use among the elderly. During my most recent visit to the Jerusalem community, in **January2017**, it appeared that there was only one sin-

gle fluent speaker left, who, for obvious reasons, no longer had any practical use for the language apart from flagging the odd phrase to younger-generation semi-speakers. Jerusalem Domari, and most likely Palestinian Domari in general, must therefore now be considered to be nearly extinct.

## 2   2. Contact languages

Given the migration route that the Dom will have taken to reach the Middle East from South Asia, it is plausible that the language was subjected to repeated and extensive contact influences. Kurdish influences on Jerusalem Domari, some of them attributable specifically to Sorani Kurdish, and some Persian items, are apparent in vocabulary, while some of the morpho-syntactic structures (such as extensive use of person affixes, and the use of a uniform synthetic marker of remote tense that is external to the person marker) align themselves with various Iranian languages. There is also a layer of Turkic loans, some of which may be attributable to Azeri varieties while others are traceable to Ottoman rule in Palestine; such items are numerous in the wordlists compiled by Seetzen and Macalister during the Ottoman period, but are much less frequent in the materials collected a century later (for a discussion of etymological sources see **Matras2012**: 426–429).

The circumstances under which speakers of Domari first came into contact with Arabic are unknown. There are some indications of a layered influence: Domari tends to retain historical /q/ in Arabic-derived words, as in *qahwa* 'coffee', *qabil* 'before', *qaddēš* 'how much', as found in the rural dialects of the West Bank (and elsewhere), whereas contemporary Jerusalem Arabic (also used by Doms when speaking Arabic) shows a glottal stop, as in *ʔahwe, ʔabl, ʔaddēš*; the word for 'now' is *hessaʕ*, while Jerusalem Arabic has *hallaʔ*. It appears that the community has been fully bilingual in Arabic and Domari at least since the early 1800s, with knowledge of Turkish having been widespread among adults during the Ottoman rule. Due to the nature of the Doms' service economy, Arabic was an essential vehicle of all professional life, whether metalwork, hawking, begging, or performance, but Domari remained the language of the household until the introduction of compulsory school education under Jordanian rule in the 1950s–60s, at which point parents ceased to pass on the language to children. By the 1990s, use of Domari was limited to a small circle of perhaps around 40–50 elderly people. Due to the multi-generational structure of households it was rare even then for conversations to be held exclusively among Domari speakers. Domari–Arabic bilingualism has always been unidirectional, with Arabic being

the language of commerce and public interactions for all Doms, and more recently also of education and media, eventually replacing Domari as a home and community language.

# 3 3. Contact-induced changes in Jerusalem Domari

As a result of ubiquitous bilingualism among all Domari speakers, Domari talk is chequered not only with expressions that derive from Arabic but also with switches into Arabic for stylistic and discourse-strategic purposes such as emphasis, direct quotes, side remarks, and so on. The structural intertwining of Domari and Arabic and the degree to which active bilingual speakers maintain a license to incorporate Arabic elements into Domari conversation pose a potential challenge to the descriptive agenda. In the following I discuss those structures that derive from Arabic, and are shared with Arabic (in the sense that they are employed by speakers both in the context of Domari conversation and in interactions in Arabic) but constitute a stable and integral part of the structural inventory of Domari without which Domari talk cannot be formed and for which there is no non-Arabic Domari alternative. All examples are taken from the Jerusalem Domari corpus described in **Matras2012**. Examples from Arabic are based on colloquial Palestinian Arabic as spoken in Jerusalem.

## 3.1 3.1 Phonology

The entire inventory of Palestinian Arabic phonemes is available in Domari; Arabic-derived words that are used in Domari conversation (whether or not they have non-Arabic substitutes) do not undergo phonological or phonetic integration, except for the application of Domari grammatical word stress on case-inflected nouns (e.g. *lambá* 'lamp.ACC', from Arabic *lámba*). The pharyngeals [ḥ] and [ʕ] are limited to Arabic-derived vocabulary. The sounds [q], [ɣ] and [ɫ] as well as [z] and [f] appear primarily in Arabic-derived vocabulary, but there is evidence that they entered the language already through contact with Turkic and Iranian languages. Less clear is the status of the pharyngealised dental consonants [dˤ, tˤ, sˤ]. These are largely confined to Arabic-derived vocabulary, but they can also be found in inherited words of Indo-Aryan stock, where they often represent original (Indo-Aryan) retroflex sounds (cf. *ḍōm* 'Dom', *pēṭ* 'belly'). An ongoing phonological innovation that is shared with Jerusalem Arabic is the simplification of the affricate [dʒ] to the fricative [ʒ] in inherited lexemes, e.g. *džami* 'I go' > *žami*. This triggers a corresponding simplification of [tʃ] to [ʃ], as in *lači*

'girl' > *laši.*

## 3.2  3.2 Morphology

Domari does not adopt productive word-derivational templates from Arabic. Arabic inflectional morphology, however, is productive with some Arabic-derived word forms, resulting in effect in a compartmentalised morphological structure. Arabic-derived plural nouns tend to retain Arabic plural inflection, but indigenous (inherited, Indo-Aryan) plural inflections are added to the word: thus *muslim* 'Muslim', plural *musilmīn-e* Muslims-PL 'muslims'; *madrase* 'school', dative plural *madāris-an-ka* schools-PL.OBL-DAT 'to the schools'. While Jerusalem Domari retains inherited plural marking with both Arabic-derived and Indo-Aryan nouns, in the closely related variety of the nomadic Doms of Jordan the Arabic plural ending *-āt* is often used with inherited nouns: thus *putur* 'son', Jerusalem Domari plural *putr-e*, Jordanian Domari plural *putr-āt.*

Arabic person agreement inflection is retained with Arabic-derived modal and aspectual auxiliaries. The auxiliaries *kān* 'be', *ṣār* 'begin', and *baqa* 'continue' take Arabic verbal inflection while *bidd-* 'want', *ḍall-* 'continue', and *xallī-* 'allow' take Arabic nominal-possessive marking:

(1)  kān-at     par-ar-m-a      wāšī-s
     was-3FSG take-3SG-1SG-PST with-3SG

     'She used to take me with her.'

(2)  dōm-e kān-u kam-k-ad-a       ḥaddādīn-e
     dom-PL was-3PL work-TR-3PL-PST blacksmiths-PL

     'The Dom used to work as blacksmiths.'

(3)  ṣār      qaft-ar-i      min ɔy-os
     began.M steal-3SG-PROG from father-3SG

     'He started to steal from his father.'

(4)  ṣār-u      kar-and-i     ḥafl-e
     began.3PL do-3PL-PROG party-PL

     'They started to have parties.'

(5)  š-ird-i     ama-ke bidd-ha      qumn-ar
     say-PFV-F 1SG-BEN want-3FSG eat-SUBJ.3SG

     'She said to me that she wants to eat.'

(6) bidd-i      par-am        itžawwiz-om-is
    want-1SG take-1SG.SUBJ marry-1SG.SUBJ-3SG.OBL
    'I want to take her and marry her.'

(7) xallī-hum       naḍḍif-k-ad-i       ehe      marn-an
    let.2SG.IMP-3PL clean-TR-3PL-PROG these.PL dead-OBL.PL
    'Let them clean up these corpses.'

(8) xallī-h rʕi-k-ar          hundar
    let-3SG graze-TR-3SG.SUBJ there
    'Let it graze there.'

Inflected Arabic-derived auxiliaries include the existential verb *kān-* 'to be', which is used in Domari, as in Arabic, as a past- and future-tense copula, supplementing the Domari remoteness or 'external' past-tense marker *-(y)a* which follows the lexical predication or predicate object:

(9) ihi    illi par-d-om-is        kān-at   yatīm-ēy-a
    this.F REL take-PST-1SG-3SG.OBL was-3FSG orphan-PRED.SG-PST
    'The one [woman] whom I married [her] was an orphan.'

Arabic-derived auxiliaries are also inflected for tense following Arabic paradigms:

(10) lāzem tkūn       itme    mišaṭṭaṭ-hr-es-i
     must  be.SUBJ.3FSG you.PL disperesed-ITR-2PL-PROG
     'You must remain dispersed.'

This amounts in effect to a functional compartmentalisation in verbal morphology, with lexical verbs, both inherited and Arabic-derived, taking inherited Indo-Aryan inflection, while Arabic-derived modal and aspectual auxiliaries take Arabic-derived inflection (for further discussion see **Matras2015**).

Arabic person agreement inflection also accompanies the Arabic-derived secondary pronominal object marker *iyyā-*, complementiser *inn-*, and conjunction *liʔann-* 'because':

(11) ple      illi t-or-im          iyyā-hum
     money.PL REL give.PST-2SG-1SG.OBL OBJ-PL
     'the money that you gave [it] to me'

(12)  aɣlabiyy-osan š-ad-i      inn-hom  min  šamāl-os-ki
      majority-3PL  say-3PL-PROG comp-3PL from north-3SG-ABL
      hnūd-an-ki
      india-OBL.PL-ABL
      'Most of them say that they are from northern India.'

(13)     na        kil-d-om barra    liʔann-hā   wars-ar-i
      NEG exit-PFV-1SG out      because-3FSG rain-3SG-PRES
      'I did not go out because it was raining.'

(14)  payy-os      liʔinn-o    ṭāṭ-i      kān
      husband-3SG because-3MSG Arab-PRED.SG was.3MSG
      'because her husband was an Arab'

Note that in example (??) the agreement is in the feminine singular, corresponding to the grammatical mapping of the Jerusalem Arabic construction 'it rains' where the (underlying) subject is the feminine noun *dunya* 'the world', while in (??) resumptive pronoun agreement with 'money', a plural noun, is in the plural.

Domari is seemingly an exception to the frequently cited generalisation that derivational morphology is more likely to be borrowed than inflectional morphology (cf. **Moravcsik1978**; **Field2002**; **Matras2009**: §6.2.2). In fact, the constraint on the borrowing of word-derivational morphology results from the clash with the principle of the transparency of morphemes (cf. **Matras2009**: §6.2.2): Arabic has few if any word-derivational morphemes that can be isolated, relying instead on complex morphological templates into which lexical roots are inserted. Nominal plural morphemes have both inflectional function (relevant to other elements in the clause) and derivational function (having independent meaning in standalone expressions). As shown above, they are replicated in Jerusalem Domari as an integral part of Arabic plural word-forms. On the other hand, the replication of inflectional material on auxiliaries is not productive in that it is not incorporated into the general lexicon, not even with lexical words of Arabic origin, but remains confined to the near wholesale adoption of modal and aspectual auxiliaries from Arabic. In this respect, Arabic-derived inflectional paradigms in Domari constitute a case of both FUSION as defined in **Matras2009** – the wholesale non-separation of language systems around a particular functional category – and at the same time a case of functional compartmentalisaton as defined in **Matras2015** – the distinct treatment of functional sub-components of a category, here the verbal category, in regard to grammatical inflection.

### 3.3   3.3 Syntax

Generally, Jerusalem Domari shows full congruence with Palestinian Arabic in most syntactic functions. This includes word order rules and the formation of both simple and complex clauses. It also includes configurations such as mapping of tenses and modality to complement and conditional clauses, and the mapping of semantic relations onto case markers. The latter can be adpositional or inflectional. For nominal possessive constructions, Domari has two options. The first of those options, illustrated in (??), is what we might call canonical Domari. It corresponds to the inherited Indo-Aryan pattern. The second option, illustrated in (??), corresponds to the common Palestinian Arabic construction, which is presented in (??). Here Domari replicate the role of the Arabic dative preposition *la* by means of the inherited Domari ablative/possessive inflectional ending *-ki*:

(15)   a.   Canonical Domari
             bɔy-im      kuri
             father-1SG house

       b.   Convergent Domari
             kury-os      bɔy-im-ki
             house-3SG father-1SG.OBL-ABL

       c.   Arabic
             bēt-o         la-ʔabū-y
             house-3SG to-father-1SG

             'my father's house'

The canonical position of adjectives in Domari is, as in other Indo-Aryan languages, before the noun (??), while in Arabic adjectives follow the noun. However, speakers show an overwhelming preference for avoiding pre-posed adjectives and instead make use of the non-verbal predication marker in order to allow the adjective to follow the noun (??), thereby replicating Arabic word order patterns (??):

(16)   Canonical Domari
       er-i             qišṭoṭ-i šōni
       come.PFV-F little-F   girl

       'A little girl arrived.'

(17)   Convergent Domari

er-i        šōni qišṭoṭ-ik
come.PFV-F girl  little-PRED.FSG
'A little girl arrived.' [= 'A girl arrived, being little.']

(18)   Arabic
ʔiža-t          bint zɣīr-e
come.PFV-F.SG girl  little-F
'A little girl arrived.'

The emergence of nominal clauses, facilitated by the availability of non-verbal predication markers, might be regarded as an innovation for an Indo-Iranian language which reinforces sentence-level convergence between Domari and Arabic:

(19)   a.   Domari
            wuda bizzot-ēk
            old.M poor-PRED.SG
       b.   Arabic
            l-xityār    miskīn
            DEF-old.M poor.M
            'The old man is poor.'

Domari, like Arabic, shows a strong tendency toward SVO word order in categorical sentences in which a thematic perspective is established by linking to a known topical entity:

(20)   mām-om putur yāsir  gar-a      swēq-ē-ta
       uncle-1SG son    Yassir go.PFV-M market-OBL.F-DAT
       'My (paternal) cousin Yassir went to the market.'

By contrast, as seen in example (??), Domari shows consistent convergence with Arabic in regard to the position of the subject after the verb when new topical entities are introduced, especially with verbs that convey movement and change of state and in presentative constructions. Drawing on inherited morphology, this convergence in word order patterns also allows for the encoding of the pronominal experiencer-recipient through a person affix that is attached to an intransitive verb in presentative constructions, matching the Arabic construction:

(21)   a.   Domari
            er-os-im              xabar
            come.PFV-3SG-1SG.OBL notice

b. Arabic

ʔaž-ā-ni            xabar

come.PFV-3MSG-1SG.OBL notice

'I received notification'

Complex clauses are also congruent with Arabic. Like Arabic, Domari shows three distinct co-temporal adverbial constructions. In the first, the subordinate clause is introduced by the conjunction 'and' and the verb is finite and indicative:

(22)  a.  Domari

kahind-ad-i    ū    pandži našy-ar-i

look-3PL-PROG and 3SG    dance-3SG-PROG

b.  Arabic

b-yi-tfarraž-u w    hiyye b-t-urʔuṣ

IND-3-look-PL and she    IND-3.FSG-dance

'They watch her dance.'

In the second, the subordinated predicate appears in the present participle:

(23)  a.  Domari

lah-erd-om-is      mindir-d-ēk

see-PFV-1SG-3SG.OBL stand-PFV-PRED.MSG

b.  Arabic

šuf-t-o        wāʔef

see.PFV-1SG-3SG standing.MSG

'I saw him standing.'

The final option shows a nominalised verb, whose possessive inflection indicates the subject/agent, introduced by the preposition 'with' in the subordinate position alongside a finite main clause:

(24)  a.  Domari

maʕ šuš-im-ki      tiknaw-ar-m-i    gurg-om

with sleep-1SG.OBL-ABL hurt-3SG-1SG-PROG neck-1SG

b.  Arabic

maʕ nōmt-i  b-t-ūžaʕ-ni     raʔbat-i

with sleep-1SG IND-3FSG-hurt-1SG neck-1SG

'As I sleep, my neck hurts.'

Relative clauses follow the format of Arabic relative clauses: They employ the Arabic-derived post-nominal relativiser *illi* and show the same distribution rules for pronominal resumption as in Arabic:

(25)  ihi    illi  par-d-om-is               kān-at     yatīm-ēy-a
      this.f REL   take-PST-1SG-3SG.OBL       was-3FSG   orphan-PRED.SG-PST
      'The one [woman] whom I married [her] was an orphan.'

Factual (indicative) complements are introduced by the Arabic-derived complelemtiser *inn-*, which carries Arabic-derived inflection (see above), and show comparable clause structure as in Arabic:

(26)  a.  Domari
          džan-ad-i        in-nā     dōm
          know-3PL-PROG    COMP-1PL  Dom
      b.  Arabic
          b-yi-ʕraf-u       in-na     dōm
          IND-3-know-3PL    comp-1PL  Dom
          'They know that we are Dom.'

Modal complements and same-subject purpose clauses show, as in Arabic, a subjunctive complement, without a complementiser:

(27)  a.  Domari
          bidd-i    dža-m        ḥaram-ka      ṣalli-k-am
          want-1SG  go-1SG.SUBJ  mosque-DAT    pray-TR-1SG.SUBJ
      b.  Arabic
          bidd-i    a-rūḥ        ʕa-l-ḥaram     a-ṣalli
          want-1SG  1SG-go.SUBJ  to-DEF-mosque  1SG-pray.SUBJ
          'I want to go to the mosque to pray.'

Adverbial clauses employ Arabic-derived adverbial subordinators such as *lamma* 'when', or composite conjunctions consisting of a preposition and complementiser, such as *baʕd mā* 'after' and *qabil mā* 'before', and generally follow Arabic sentence organisation and tense and modality distribution patterns:

(28)  lamma lak-ed-a xāl-os       indžann-ahr-a        bɔy-om
      when  see-PFV-M uncle-3SG   crazy-TR.PFV-M       father-1SG
      'When he saw his uncle, my father went crazy.'

(29)  baʕd mā    xaḷḷaṣ-k-ed-a   kam-os    gar-a      kury-is-ta
      after COMP finish-TR-PFV-M work-3SG go.PFV-M house-3SG.OBL-DAT
      'After he finished his work he went home.'

(30)  qabil mā    dža-m         xaḷḷaṣ-k-ed-om    kam-as
      before COMP go-1SG.SUBJ finish-TR-PFV-1SG work-OBL.M
      'Before I left I finished my work.'

Conditional clauses similarly draw on Arabic the conjunctions *iza* and *law*, both 'if', and show similar distribution of tense and aspect categories, including the Arabic-derived impersonal marker of counter-factuality *kān,* literally 'was':

(31)   a.  Domari
           law er-om        xužoti    kān lah-erd-om-s-a
           if  come.PFV-1SG yesterday was see-PFV-1SG-3SG-PST
       b.  Arabic
           law žī-t         mbāreḥ   kān šuf-t-o
           if  come.PFV-1SG yesterday was see.PFV-1SG-3SG
           'If I had come yesterday, I would have seen him.'

## 3.4   3.4 Lexicon

Jerusalem Domari shows extensive impact of Arabic on the grammatical lexicon, including almost wholesale reliance on Arabic-derived material for entire categories. In the pronominal domain, Domari employs, in additional to the secondary pronominal object marker *iyyā-* discussed above, also the Arabic reflexive pronoun *ḥāl-*, derived from the word 'state', in combination with person/possessive inflection, and the Arabic reciprocal pronoun *baʕḍ-*:

(32)  naḍḍif-k-ad-a   ḥāl-os
      clean-TR-PFV-M REFL-3SG
      'He cleaned himself.'

(33)  tʕarraf-h-r-ēn    baʕḍ-ē-man-ta
      meet-TR.PFV-1PL RECP-PL-1PL-DAT
      'We met one another.'

Indefinite expressions draw on Arabic-derived forms of category determination including negative *wala*, free choice *ayy* and universal *kull,* which may be

combined with inherited ontological markers, as well as on the ontological specifiers *ḥādž-* for thing and *maḥall* for location. Indefinite expressions that derive entirely from Arabic include temporal *wala marra* 'never', dāyman 'always', and universal-thing kullši 'everything'. Arabic-derived focus particles are *barḍo* 'also, too' and *ḥatta* 'even' and quantifiers are *kull* 'every, each' and *akamm* 'a few'. Interrogatives are generally inherited (Indo-Aryan), with the exception of *qaddēš* 'how much'. Numerals are all derived from Arabic with the exception of the lowest numeral forms (1–5 in citation function and 1–3 in attributive role) (see Tables 1–2); all ordinal numerals (*awwal* 'first', *tānī* etc.) are from Arabic.

| Numeral | Citation | Attribute |
|---|---|---|
| 1 | *ikak* | *-ak* |
| 2 | *diyyes* | *di* |
| 3 | *taranes* | *taran* |
| 4 | *štares* | *ʔarbaʕ* |
| 5 | *pʌndžes* | *xamis* |
| 6 | *sitt-ēk-i* | *sitt* |
| 7 | *sabʕ-ak-i* | *sabaʕ* |
| 8 | *tamāni-ak-i* | *tamānye* |
| 9 | *tisʕ-ak-i* | *tisʕa* |
| 10 | *das* 'ten', *ʕašr-ak-i* | *ʕašr* |
| 20 | *ʕišrīn-i, wīs-i* | *ʕišrīn* |
| 21 | *ʕišrīn ū ekak-i* | *wāḥed w ʕišrīn* |
| 22 | *ʕišrīn-i ū diyyes-i* | *tnēn w ʕišrīn* |
| 23 | *ʕišrīn-i ū taranes-i* | *talāte w ʕišrīn* |
| 24 | *ʔarbaʕ ū ʕišrīn* | *ʔarbaʕ w ʕišrīn* |
| 100 | *miyyēk hi, siyy-ak-i* | *miyye* |
| 1000 | *alf-ak-i* | *alf* |

*Table ??: Jerusalem Domari numerals*

| Numeral | Form |
|---|---|
| 30 | *talātīn* |
| 40 | *ʔarbaʕīn* |
| 50 | *xamsīn* |
| 60 | *sittīn* |
| 70 | *sabʕīn* |
| 80 | *tamanīn* |
| 90 | *tisʕīn* |

*Table ??: Jerusalem Domari higher numerals*

Alongside a very small number of inherited prepositions that are used exclusively with pronominal (person-inflected) forms, most prepositions are derived from Arabic (Table ??).

| | | | | | |
|---|---|---|---|---|---|
| *ʕan* | 'on, about' | *ʕašān* | 'because' | *nawāḥi* | 'toward' |
| *maʕ* | 'with' | *minšān* | 'for' | *qabil* | 'before' |
| *min* | 'from' | *min ɣēr* | 'without' | *baʕd* | 'after' |
| *la, ʕala* | 'to' | *bidūn, min dūn* | 'without' | *laɣāyet* | 'until' |
| *fi* | 'in' | *bēn* | 'between' | *bi* | 'in, for' |
| *zayy* | 'like' | *ḥawāli* | 'around' | *ḍiḍ* | 'against' |
| *ʕind (ʕand)* | 'at' | *badāl* | 'instead of' | *min ḍamn* | 'among' |
| *žamb* | 'next to' | *ʔilla ɣēr* | 'except for' | | |

*Table ??: Arabic-derived prepositions in Jerusalem Domari*

Arabic-derived grammatical operators at verbal clause level include a series of modality adverbs such as *masalan* 'for example', *yimken* 'perhaps', *atāri* 'well', time adverbs such as *hessaʕ* 'now' and *baʕdēn* 'then, afterwards', and the phasal adverbs *lissa* and *lāyzāl*, both 'still'. As discussed above, Domari adopts Arabic modal and aspectual auxiliaries wholesale, i.e. along with their Arabic-derived inflection. This covers almost the full category of modal and aspectual auxiliaries including habitual/iterative *kān* 'be', *ṣār* 'begin', and *baqa* 'continue', *bidd-* 'want', *ḍall-* 'continue', and *xallī-* 'let', as well as the impersonal form *lāzem* 'must'. The only modal for which an Indo-Aryan form is retained is *sak-* 'to be able to'. Past-tense finite predications take the Arabic negator *mā* (alongside inherited *na*) while in non-finite predications the Arabic negation particle *miš* is used:

(34)  mā  lak-ed-om-is
      NEG see-PFV-1SG-3SG.OBL
      'I didn't see him/her.'

(35)  bay-os   miš kury-a-m-ēk
      wife-3SG NEG house-OBL.F-LOC-PRED.SG
      'His wife is not at home.'

Clause combining relies exclusively on Arabic-derived material (connectors and conjunctions) (see Table ??).

| w | 'and' | qabil mā | 'before' |
|---|---|---|---|
| wala | 'and not', 'neither', 'either' | baʕd mā | 'after' |
| yā | 'or' | min-yōm-mā | 'since' |
| willa | 'or', 'or else', 'neither' | iza | 'if' |
| bass | 'but', 'only', 'however' | law | 'if' |
| illi | relative pronoun | bi-r-raɣim | 'despite', 'although' |
| ʔinn- | 'that' | ʕašān | 'for', 'in order to' |
| liʔann | 'because' | minšān | 'for', 'in order to' |
| lamma | 'when' | ta | 'in order to' |
| kull mā | 'whenever' | | |

*Table ??: Arabic-derived conjunctions in Jerusalem Domari*

Likewise, the inventory of discourse particles and interjections is adopted wholesale from Arabic: We find the interjection, tags and filers *yabayyi, yaḷḷa, xaḷaṣ, waḷḷa,* and *yaʕni,* as well as segmental markers with a lexical meaning such as *l-muhimm* 'anyway', *l-ḥāṣil* 'finally', *ṭayyib* 'well', *w ʔiši* 'and the like', *w hāda* 'and so on', *abṣar* 'whatever', and the filler *hāy* 'that'. The quotation particle *qal/xal,* from Arabic 'say', is not found in Jerusalem Arabic and appears to represent an older layer of Arabic influence (as indicated also by its phonological structure; see introductory remarks).

The content lexicon equally shows massive impact of Arabic. In the Jerusalem Domari corpus of narrational and conversational talk as well as sentence elicitation recorded in the 1990s (**Matras2012**) almost two thirds of lexical items are Arabic-derived; the count includes single word insertions from Arabic, including attributive nominal compounds (noun–possessor and noun–adjective), but excludes phrases containing a finite lexical verb that is Arabic-derived (which latter are regarded as optional code-switches). Both Arabic-derived nouns and adverbs outnumber inherited (Indo-Aryan) counterparts by around 65% to 35%, while for verbs and adjectives the numbers are roughly equal. Around 26% of items of both the Swadesh 100-item list and the Leipzig–Jakarta 100-item list (**HaspelmathTadmor2009**) are Arabic-derived. This puts Domari in the range of languages considered to be 'high borrowers' by the Leipzig Loanword Typology Project (**HaspelmathTadmor2009**). Meanings on the list that are replaced by Arabic loans in Domari include a number of animals ('ant', 'bird', 'fish'), activities

('to run', 'to fly', 'to crush'), elements of nature ('star', 'soil', 'shade', 'ash', 'leaf', 'root'), and some body parts ('knee', 'navel', 'liver', 'thigh'; also 'wing', 'tail'). On the whole, the meaning and usage of Arabic-derived lexemes matches that of Jerusalem Arabic. Creative processes are marginal and include such processes as the phonological volatility of /q/ (as /q/, /x/, /qx/ and /g/), the alternation between *fardžik-* 'to show' (Arabic *f.r.dž*) and *wardžik-*, and the occasional creative derivation such as *bisawahr-* 'to get married', from Arabic *bi-sawa* 'together'.

Arabic verbs are integrated into Domari through a light verb construction that draws on the inherited verb stems *-k-* 'to do' and *-h-* 'to become', which are grammaticalised into loan verb adaptation markers (see **Matras2012**: 240–244) that are sensitive to valency. This follows a strategy for the adaptation of loan verbs that is widespread across a geographical area stretching from the Balkans and the Caucasus through Anatolia and Western Asia and on to the Indian Subcontinent. For some verbs, alternating adaptation markers can indicate change in valency: *džawwiz-h-r-i* marry-ITR-PFV-F 'she got married', *džawwiz-k-am-is* marry-TR-1SG.SUBJ-3SG.OBL 'I shall marry her off'. The core of integrated Arabic verbs generally derives from the Arabic subjunctive–imperative form, which in Arabic never occurs in isolation from its person inflection in the prefix conjugation, as in *džawwiz-* 'marry', from *\*džawwiz!* 'marry (off)!' or *\*t-džawwiz!* 'get married!'. Note however that the vowel structure of the core does not always correspond to the subjunctive–imperative form of contemporary Palestinian Arabic, which is quite possibly a further indication of the layered historical influence of Arabic. Thus we find *s'il-k-ed-om* ask-TR-PFV-1SG 'I asked', from *\*s'il-* 'ask', while Palestinian Arabic has *isʔal!* 'ask!', and *rawwaḥ-ah-r-a* go-ITR-PFV-M 'he travelled', while Palestinian Arabic has *rawwiḥ!* 'go away!'.

## 3.5    3.5 Cross-category interplay

A typologically curious case of contact-induced change is offered by the use in Jerusalem Domari of three construction types that cut across structural categories. The first pertains to the comparative form of adjectives. In the absence of a structurally transparent, isolated and replicable marker of adjective comparison (comparative and superlative), Domari draws on Arabic word-forms for all comparative adjective forms, even when an inherited (non-Arabic) word form is used for the positive form of the adjective:

(36)   Domari

atu      qaštot-ik
you.SG small-PRED.FSG
'You are small.'

(37)  Domari
      atu      azɣar  mēšī-m-i
      you.SG smaller from-1SG-PRED.SG
      'You are smaller than I.'

(38)  Arabic
      inti     zɣīr-e
      you.FSG small-FSG
      'You are small.'

(39)  Arabic
      inti     azɣar  minn-i
      you.FSG smaller from-1SG
      'You are smaller than I.'

This formation involves essentially the recruitment of an alternative, Arabic-derived item from the category of lexical items in order to carry out a grammatical procedure that is derivational–inflectional by nature (derivational in that it modifies meaning, inflectional in that it is inherently embedded into a syntactic relationship at the phrase level); thus we have a case of cross-category interplay.

A further case is that of lexical suppletion around Arabic-derived numerals. Domari and Arabic differ typologically in respect of numeral agreement: with Indo-Aryan numerals, the Domari noun appears in the default singular form, while in Arabic, numerals up to 10 take plural agreement. The clash is resolved in Domari in such a way that Arabic-derived numerals under 10 invariably trigger an Arabic-derived lexical item even when an inherited form of the corresponding lexeme is available:

(40)  ḥkum-ke-d-os            taran wars, maḥkame
      sentence-TR-PFV-3SG three year   court
      'The court sentenced him to three years.'

(41)  eh-r-a               ʕumr-om sitte snīn
      become-PFV-M age-1SG   six   year.PL
      'I turned six years old.'

Such alternation is systematic (see further examples in Table **??**) and might be regarded as a case of bilingual suppletion, where every countable noun in the language for which an inherited (Indo-Aryan) word form exists also has an Arabic-derived counterpart that is used with numerals between 3 and 10.

| Inherited numeral, inherited singular noun | Arabic numeral, Arabic plural noun |
|---|---|
| *di dīs taran dīs* 'two days three days' | *sabaʕ-t-iyyām* 'seven days' |
| *taran mas* 'three months' | *xamas-t-ušhur* 'five months' |
| *taran wars* 'three years' | *sitte snīn* 'six years' |
| *taran zard* 'three pounds' | *xamas līrāt* 'five pounds' |

*Table ??: Some phrases from the corpus containing numerals and nouns*

Finally, while Domari lacks a definite article, the Arabic definite article *l-* is employed with definite noun phrases where both the noun and the numeral-attribute are derived from Arabic:

(42) mar-d-e      l-ʔarbaʕ xurfān
     kill-PFV-3PL DEF-four lambs.PL

     'They slaughtered the four lambs.'

(43) dīr-os       it-tānye      eh-r-i        muhandis-ēk
     daughter-3SG DEF-second.F become-PFV-F engineer-PRED.FSG

     'Her other daughter became an engineer.'

# 4   4. Conclusion

The comparison with **Macalister1914**'s (**Macalister1914**) materials offers some scope for observations in respect of the historical development of contact-induced change over the past century in at least two areas of structure, namely the loss of Turkish-derived vocabulary as well as of some of the inherited Indo-Aryan vocabulary (around 55 words are attested in Macalister's materials that were not familiar to the speakers I interviewed), and the adoption of fully inflected modal and aspectual auxiliaries, compared to their use as impersonal forms in Macalister's material. One has to bear in mind, however, that Macalister's corpus is based on work with just a single speaker. Nevertheless, these changes provide some indication that the impact of Arabic continued to expand during the last century in which the language was spoken, a period during which the Doms lost much of their distinct culture and lifestyle as a result of the shift from a semi-

nomadic service economy to a settled, wage-based but still socially isolated and stigmatised community.

The impact of Arabic on Domari prompts a theoretical challenge around identifying a form of the language that is structurally inseparable from Arabic. This can be illustrated by the following two examples:

(44)  a. Domari
         aktar min  talātīn xamsa w   talātīn sana mā  lak-ed-om-is
         more from thirty  five     and thirty  year NEG see-PFV-1SG-3SG.OBL
      b. Arabic
         aktar min  talātīn xamsa w   talatīn sana mā  šuf-t-hā
         more from thirty  five     and thirty  year NEG see.PFV-1SG-3FSG

         'I haven't seen her for more than thirty, thirty five years.'

(45)  a. Domari
         kān       ʕumr-om yimken sitte snīn   sabʕa snīn
         was.3MSG age-1SG   maybe  six   years seven years
      b. Arabic
         kān       ʕumr-i yimken sitte snīn   sabʕa snīn
         was.3MSG age-1SG maybe   six   years seven years

         'I was maybe six or seven years old.'

Both (??) and (??) are unambiguously identifiable to speakers as Domari utterances; moreover, their meaning cannot be conveyed in Domari in any other way. Yet they each differ in just one single element from their respective counterpart Arabic utterances in (??) and (??): the use of the lexical verb with subject and object agreement (Domari *lak-ed-om-is* 'I saw her', Arabic *šuf-t-ha*) in the first, and the use of the 1.SG possessive marker (Domari *-om*, Arabic *-i*) in the word *ʕumr* 'age' in the second. Despite being isolated examples, (??)–(??) illustrate the considerable extent of structural overlap between the two languages. Furthermore, the examples discussed above of bilingual suppletion in number agreement and adjective comparison, and the productive use of Arabic person agreement inflection with auxiliaries and with some complementisers and secondary object markers, mean in effect that active command of Arabic is a pre-requisite for speaking Domari.

It follows that Domari provides us with an opportunity to reconsider the taxonomy of contact-induced language change phenomena. It is not a Mixed Language by conventional definitions (cf. **BakkerMatras2013**; **Matras2009**: chapter 10) since the Indo-Aryan source of grammatical inflection in all word classes is

overwhelmingly consistent with the source of basic lexical vocabulary and of deictic and anaphoric elements (demonstrative and personal pronouns, interrogatives, and spatial adverbs). Impressionistically speaking, it is a language with 'heavy borrowing' in that it shows the adoption of Arabic-derived material in a wide range of different structural categories. But the distribution of some of this material, taking into account the ubiquitous active bilingualism among Domari speakers, lends itself to the postulation of several particular types of contact-induced structural change, which I have labeled above FUSION (wholesale non-separation of languages around a particular structural category, e.g. clause connectors and modal auxiliaries), inflectional compartmentalisaton (the use of Arabic inflectional paradigms with particular functional categories, notably modal and aspectual auxiliaries), and bilingual suppletion (activation of speakers' full command of Arabic vocabulary and inflection for creative formations around number agreement and adjective comparison).

## 5 Further Reading

**Matras2007** outlines contact influences on Jerusalem Domari in the context of a collection of chapters on contact-induced change in a sample of different languages. **Matras2012** provides a descriptive and historical overview of Jerusalem Domari and includes extensive discussion of contact-induced change in the individual chapters as well as a chapter devoted to the impact of Arabic. **Matras2009** is a general theoretical discussion of contact-induced change in functional-typological perspective and includes many examples from Jerusalem Domari. Finally, **Matras2015** discusses patterns of morphological borrowing and their theoretical implications and gives as one of the examples the compartmentalisaton of modal and aspectual auxiliaries in Jerusalem Domari.

## 6 Abbreviations

ABL ablative
  BEN benefactive
  COMP complementiser
  DAT dative
  F feminine
  IMP imperative
  IND indicative
  ITR intransitive

LOC locative
M masculine
OBL oblique
PFV perfective
PL plural
PRED predication (non-verbal)
PRES present
PROG progressive
PST past
RECP reciprocal
REFL reflexive
REL relativiser
SG singular
SUBJ subjunctive
TR transitive

```
%%move bib entries to  localbibliography.bib




@book{Bakker2013,
address = {Berlin},
booktitle = {\textit{The} \textit{mixed} \textit{language} \textit{debate:} \texti
editor = {Bakker, Peter \ and  Yaron Matras},
publisher = {Mouton de Gruyter},
title = {\textit{The} \textit{mixed} \textit{language} \textit{debate:} \textit{The
year = {2013}
}




@article{Black1913,
author = {Black, George Fraser},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Gypsy} \textit{Lore} \
pages = {327–330},
title = {The gypsies of {Armenia}},
volume = {6},
year = {1913}
```

```
}


@book{Field2002,
address = {Amsterdam},
author = {Field, Frederick.},
publisher = {John Benjamins},
title = {\textit{Linguistic} \textit{borrowing} \textit{in} \textit{bilingual} \
year = {2002}
}


@book{Haspelmath2009,
address = {Berlin},
author = {Haspelmath, Martin  and  Uri Tadmor.},
publisher = {De Gruyter Mouton},
title = {\textit{Loanwords} \textit{in} \textit{the} \textit{World's} \textit{La
year = {2009}
}


@article{Herin2012,
author = {Herin, Bruno},
journal = {\textit{Linguistic} \textit{Discovery}},
number = {2},
pages = {1–52},
title = {The Domari language of Aleppo ({Syria})},
volume = {10},
year = {2012}
}


@article{Herin2014,
author = {Herin, Bruno},
```

```
journal = {\textit{Zeitschrift} \textit{der} \textit{Deutschen} \textit{Morgenländ:
pages = {407--450},
title = {The northern dialects of Domari},
volume = {164},
year = {2014}
}




@article{Herin2016,
author = {Herin, Bruno},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}},
pages = {33--74},
title = {Elements of Domari dialectology},
volume = {23},
year = {2016}
}




@book{Kruse1854,
address = {Berlin},
author = {Kruse, Friedrich.},
publisher = {Reimer},
title = {\textit{Ulrich} \textit{Jasper} \textit{Seetzen's} \textit{Reisen} \texti
year = {1854}
}




@book{Macalister1914,
address = {London},
author = {Macalister, Robert A. S.},
number = {3},
publisher = {Edinburgh University Press},
series = {Gypsy Lore Society Monographs},
title = {\textit{The} \textit{language} \textit{of} \textit{the} \textit{{Nawar}} \
year = {1914}
```

}

```
@article{Matras1999,
author = {Matras, Yaron},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}},
pages = {1–58},
title = {The state of present-day Domari in Jerusalem},
volume = {11},
year = {1999}
}
```

```
@article{Matras2000,
author = {Matras, Yaron},
journal = {\textit{Romani} \textit{Studies}},
pages = {53–79},
title = {Two Domari legends about the origin of the Doms},
volume = {10},
year = {2000}
}
```

```
@incollection{Matras2007,
address = {Berlin},
author = {Matras, Yaron},
booktitle = {\textit{Grammatical} \textit{Borrowing} \textit{in} \textit{Cross-L
editor = {Yaron Matras \& Jeanette Sakel},
pages = {151–164},
publisher = {Mouton de Gruyter},
title = {Grammatical borrowing in Domari},
year = {2007}
}
```

```
@book{Matras2009,
address = {Cambridge},
author = {Matras, Yaron.},
publisher = {Cambridge University Press},
title = {\textit{Language} \textit{Contact}},
year = {2009}
}




@incollection{Matras2011,
address = {Moscow},
author = {Matras, Yaron},
booktitle = {\textit{Languages} \textit{of} \textit{the} \textit{World:} \textit{N
editor = {Tatiana I. Oranskaia, Yulia V. Mazurova, Andrej A. Kibrik, Leonid I. Kul
pages = {775-811},
publisher = {Academia},
title = {Domari},
year = {2011}
}




@book{Matras2012,
address = {Berlin},
author = {Matras, Yaron.},
publisher = {De Gruyter Mouton},
title = {\textit{A} \textit{grammar} \textit{of} \textit{Domari}},
year = {2012}
}




@incollection{Matras2015,
address = {Berlin},
author = {Matras, Yaron},
booktitle = {\textit{Borrowed} \textit{morphology},},
```

```
editor = {Francesco Gardani, Peter Arkadiev \& Nino Amridze},
pages = {47–80},
publisher = {Mouton de Gruyter},
title = {Why morphological borrowing is dispreferred},
year = {2015}
}


@incollection{Moravcsik1978,
address = {Stanford},
author = {Moravcsik, Edith A},
booktitle = {\textit{Universals} \textit{of} \textit{human} \textit{language}, v
editor = {Joseph H. Greenberg, Charles A. Ferguson \& Edith A. Moravcsik},
pages = {94–122},
publisher = {Stanford University Press},
title = {Universals of language contact},
year = {1978}
}


@article{Newbold1856,
author = {Newbold, F. R. S},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Royal} \textit{Asia
pages = {285–312},
title = {The Gypsies of {Egypt}},
volume = {16},
year = {1856}
}


@book{Paspati1870,
address = {Osnabrück},
author = {Paspati, Alexandre.},
publisher = {Biblio},
title = {\textit{Études} \textit{sur} \textit{les} \textit{Tschinghianés} \texti
```

```
year = {1870}
}




@article{Patkanoff1907,
author = {Patkanoff, K. P},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Gypsy} \textit{Lore} \
pages = {229–257},
title = {Some words on the dialects of the Transcaucasian Gypsies},
volume = {1},
year = {1907}
}




@article{Pott1846,
author = {Pott, August F},
journal = {\textit{Zeitschrift} \textit{für} \textit{die} \textit{Wissenschaft} \t
pages = {175–186},
title = {Über die {Sprache} der Zigeuner in Syrien},
volume = {1},
year = {1846}
}
```

## Acknowledgements

**Chapter 6**

# Contact-induced grammaticalization between Arabic dialects

Thomas Leddy-Cecere

Bennington College

This chapter describes the phenomenon of contact-induced grammaticalization between Arabic dialects and its significance in accounting for the development of future tense markers across modern Arabic varieties. After an introduction to theoretical aspects of general grammaticalization theory and contact-induced grammaticalization in particular, discussion shifts to the identification of specific contact-induced grammaticalization processes leading to the modern distribution of future tense-marking forms across the Arabic-speaking world. Finally, the significance of these findings to broader inquiry in Arabic dialectology and theoretical contact linguistics is considered.

## 1 Introduction & theory

### 1.1 Overview

This chapter presents evidence for the occurrence of contact-induced grammaticalization processes between dialects of Arabic over the course of the language's history. The critical role of dialect contact as a source of synchronic variation and diachronic change across Arabic varieties is well recognized, and the description of its outcomes a long-standing occupation of Arabic dialectologists (e.g., **Behnstedt2005**; **Miller2007**). Representing a fairly recent theoretical development in the field of contact linguistics – following largely from the proposals of Heine and Kuteva (**Heine2003**; **Heine2005**) and **Dahl2001** – contact-induced grammaticalization as a model has not been applied to the analysis of Arabic dialect data on a large scale. As will be seen, however, the phenomenon displays

significant merit as an account for the evolution and diffusion of a number of morphosyntactic features across the modern Arabic dialects.

In the following subsections, I begin with a review of the current state of research in grammaticalization theory and in contact-induced grammaticalization (CIG) specifically. I then proceed to an illustrative example of CIG in the Arabic context, demonstrating the model's power as an explanatory mechanism in interpreting the distribution of future tense markers across the modern dialects. I conclude the chapter with a brief discussion of the broader significance of CIG in the analysis of Arabic and the potential role for Arabic data in advancing general theoretical knowledge of the phenomenon at large.

## 1.2 Grammaticalization

Most linguists agree that it is possible to synchronically classify the majority of linguistic forms along a cline from "more lexical" to "more grammatical", in a manner roughly consistent with the progression as conceived by **Hopper2003**:

CONTENT WORD > GRAMMATICAL WORD > CLITIC > INFLECTIONAL AFFIX

Historical linguists would add to this synchronic observation the diachronic reflection that it is common to observe a single etymological item advancing through the successive stages of this cline as it develops as part of a linguistic system over time. In fact, the sheer frequency of examples indicating such a trajectory of evolution has led to the identification of a cross-linguistically attested phenomenon known as grammaticalization. The recent definition provided by Hopper and Traugott is indicative of several currently referenced in the field, which – though differing in emphasis and points of detail – broadly subscribe to a similar central principle:

> [Grammaticalization is] the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions (**Hopper2003**: 18).

Though useful for purposes of general definition, this largely intuitive formulation of grammaticalization and the cline which it follows must be further deconstructed if they are to be operationalized as part of a rigorous analysis. **Andersen2008** summarizes the issue succinctly in observing that the grammaticalization cline so articulated conflates numerous discrete dimensions of language change by presenting them as unified steps in a chain: the shift from lexical

to grammatical word is one of semantic content, while that from word to clitic to affix involves morphosyntax and any associated loss of phonological material is best understood as phonological change. Since the early stages of grammaticalization research, more complex approaches to the description of the phenomenon have been proposed based on the concurrent evaluation of multiple parameters (e.g., **Lehmann1985**). Other authors opt instead to define analogous parameters in terms of diachronic processes, thereby rendering them more directly relatable to the modes of historical linguistic analysis which underlie the bulk of investigations in grammaticalization research. The latter approach is adopted here, largely following the account proposed by **Heine2007**.

Heine views grammaticalization as defined by the simultaneous progression of four distinct but interrelated diachronic processes: desemanticization, extension, decategorialization, and erosion. Desemanticization involves the loss of concrete lexical ("content") meaning and a corresponding rise in abstract grammatical function associated with the use of an item in particular contexts. This often represents the first observable stage of grammaticalizing change, and, as its name suggests, primarily concerns the semantic content of the item rather than its distribution, form, or syntactic behavior. Closely coupled with desemanticization is extension, namely the novel use of the grammaticalizing item in contexts in which it was not previously employed; extension is thus defined as a change in incidence. The hand-in-hand advance of these two processes is demonstrated in the evolution of the French negative element *pas*: having shed its content semantics as a noun meaning 'step' and developed grammatical function as a marker of verbal negation, *pas* is extended in contemporary usage to contexts involving none of the implied motion of its lexical source (**Hansen2009**: 137–138).

The third process described by Heine, that of decategorialization, consists of the changes by which a grammaticalized item comes to lose the morphosyntactic properties characteristic of its source's original word class, such as word order freedom or agreement inflection; an example may be found in the gradual development of the English adverbial marker *-ly* from a morphosyntactically free substantive meaning 'body, form' to a bound derivational suffix (**Ramat2011**). Erosion, the fourth process considered by Heine, refers to the gradual reduction and lenition of phonological form beyond what is accounted for by regular sound change, as observed in the irregular changes deriving the Jewish Babylonian Aramaic continuous aspect marker *qā ~ kā* from earlier *qāʔē 'standing' (**Rubin2005**).

Theories of grammaticalization have also been strongly linked to the notion of unidirectionality, the proposal that change along the above-described cline occurs only from more lexical to more grammatical and not vice versa (**Lehmann2015**).

Though the absolute formulation of this hypothesis has been the subject of much debate (e.g., **Norde2009**), recognition of a strong unidirectional tendency remains integral to understandings of grammaticalization on both empirical and theoretical grounds (**Haspelmath1998**; **Heine2007**). It has been proposed that the impetus for such a tendency lies in a universal set of cognitive and communicative principles common to the human mental faculty (**Claudi1986**; **Bybee2003**; **Lehmann2015**); these would provide an account for the pervasive occurrence of grammaticalization as a worldwide phenomenon, and may be seen to bias the results of grammatical change in the directions entailed by the four processes described above.

The concomitant advancement of these processes is discernible in one of the few cases of Arabic grammaticalization for which a reasonably complete chain of historical development is attested: that of the Egyptian Arabic future tense marker *ḥa-*. Documented in 16th and 17th century sources as *rāyiḥ*, this item already shows evidence of desemanticization and extension, departing from the semantics of its lexical source as an active participle 'going' to indicate intention and imminent futurity of action, consequently allowing its extension to usage contexts devoid of actual motion: *ʔanā rāyiḥ aɣannī ʕalēh* 'I am going to sing about it (and proceeds to sing)' (**Davies1981**). In its 19th century incarnation *rāḥ ~ raḥ ~ ḥa-*, the form shows further desemanticization and extension as its value changes from an imminent to a general future and it comes to be employed in previously unacceptable circumstances, such as in the presence of a non-immediate temporal adverb: *rāḥ yīgi bukra* 'he'll come tomorrow' (**Elias1981**: 157; for earlier usage constraints, see **Davies1981**: 241-243). These increasingly modern forms also attest decategorialization, as the once-obligatory adjectival agreement marking of the participial original becomes optional – *raḥ* (MSG)/*raḥa* (FSG)/*raḥīn* (PL) ~ *raḥ* (invar.) (**Vollers1895**) – and eventually ceases to exist altogether in the tightly bound modern clitic *ḥa- ~ ha-* (**Abdel-Massih2009**: 268). Fourthly, progressive phonetic erosion is visible throughout the item's history, as none of the loss or lenition of phonetic material through the stages *rāyiḥ > rāḥ > raḥ > ḥa- > ha-* attested above is attributable to regular sound change. Taken together, these combined processes chart the grammaticalization of lexical *rāyiḥ* 'going' through its gradual development into the modern future tense clitic *ḥa-*.[1]

Having established these understandings of grammaticalization and its com-

---

[1] On sources referenced in the preceding paragraph: **Davies1981** is a study of colloquial elements in the seventeenth century Egyptian text *Hazz al-quḥūf fī šarḥ qaṣīd Abī Ṣādūf*; **Vollers1895** is a descriptive grammar of Egyptian Arabic at the close of the 19th century, and **Elias1981** an English–Egyptian Arabic vocabulary and phrasebook first released ca. 1899; **Abdel-Massih2009** is a reference grammar of modern Egyptian Arabic.

ponent processes, we now turn to the proposal that specific grammaticalization pathways are able to be shared between interacting languages and dialects: the aforementioned CIG.

## 1.3   Contact-induced grammaticalization

The most elaborated theoretical model proposed for CIG is that of Heine and Kuteva **Heine2003**; **Heine2005**. This model represents the phenomenon by which "a grammaticalization process … is transferred from the model (M) to the replica language (R)," without corresponding transfer of any actual phonological form (2003: 539). As paraphrased and clarified by **Law2014**, this occurs when one language, "the 'replica language,' develops a feature observed in another language, the 'model' language, but goes through a path of universal development using resources internal to the replica language." The result is such as that seen in the Basque innovation of an allative preposition from the noun *buru* 'head' and a perfect tense formed with the lexical verb *ukan* 'have', apparently influenced by parallel grammaticalizations in neighboring varieties of Romance (**Haase1992** *apud* **Heine2003**: 550). Such an effect is proposed to be actuated according to the following model:

1. Speakers of language R notice that in language M there is a grammatical category $M_x$.

2. They develop an equivalent category, $R_x$, using material available in their own language (R).

3. To this end, they replicate a grammaticalization process they assume to have taken place in language M, using an analogical formula of the kind $[M_y > M_x] = [R_y > R_x]$.

4. They grammaticalize $R_y$ to $R_x$. (**Heine2003**: 539)

This proposal for the diffusion of parallel grammaticalization trajectories across linguistic varieties is presaged by **Bisang1998**'s (**Bisang1998**) observation of the potential synergy between grammaticalization, which he views primarily as a construction-based process, and previously observed forms of contact-induced structural convergence. The phenomenon has also been influentially described by Dahl in the form of "gram families," consisting of groups of "grams [grammaticalized items] with related functions and diachronic sources that show up in genetically and/or geographically related groups of languages, in other words,

what can be assumed to be the result of one process of diffusion" (**Dahl2001**: 1469). Heine and Kuteva draw heavily on Dahl's theorizations, though they diverge from him in a few critical ways. First, they are significantly more conservative than Dahl in identifying examples of the phenomenon, insisting on corroborating evidence of language contact in order to posit CIG rather than inductively inferring its occurrence given genetic relatedness or proximity. Second, they do not necessarily attempt to link multiple replications of the same grammaticalization pathway into "one process of diffusion," but instead prefer to treat them as individual instances of contact between participating languages.

Further, Heine and Kuteva's model is primarily situated in the context of contact between genetically distinct languages. Regarding the occurrence of CIG between related language varieties or dialects, Dahl sees such scenarios as generating the bulk of evidence for the phenomenon: "in the majority of all such cases [of areally diffused grams], the languages involved are more or less closely related" (**Dahl2001**: 1469). Heine and Kuteva are wary of such identifications. Critically, however, their reasons for being so are methodological rather than theoretical. In their analysis, they choose to rely on the principle of genetic patterning as "an empirically well-founded tool for identifying cases of contact-induced linguistic transfer" (**Heine2005**: 33–34), meaning that examples of CIG between unrelated languages are often easiest to identify and defend and thus have been favored in the effort to present an unambiguous account. Regarding the broader occurrence of the phenomenon, however, they state that "genetic relationship is entirely irrelevant" (**Heine2005**: 184) and that CIG may occur between related languages just as it does between unrelated ones. They remain, though, more careful than Dahl to set apart cases attributable to inheritance of any stage of the grammaticalization chain from a common ancestor, which could lead to a superficially similar result not in fact dependent on any degree of contact. Along the same lines, **Law2014** reminds us that when dealing with closely related languages the possibility of drift or typological poise precipitating parallel development rises dramatically in likelihood. Thus, the analyst must be stringent in linking proposed cases of CIG to cross-linguistically attested paths and parameters of grammaticalization and not to the local idiosyncrasies of a given language family or subgroup.

To Heine and Kuteva, CIG is unambiguously situated in terms of Van Coetsem's (**Coetsem1988**; **Coetsem2000**) dichotomy between source language (SL) and recipient language (RL) agentivity: the four-stage model of replication presented above clearly identifies speakers of the RL as the agents of contact-induced change in this instance. This judgment has opened the proposal to major critique,

as several key theorists maintain that structural pattern replication of the kind required for CIG is only possible in a scenario of SL agentivity. **Ross2007**, for example, views the phenomenon as part of a broader process of bilingual calquing, involving the subconscious imposition of the functional range of an SL item onto its RL equivalent, followed secondarily by the processes Heine and Kuteva attribute to grammaticalization but which Ross views as the natural result of increases in frequency and automization stemming from the RL item's new functionality. Ross asserts that "one cannot reasonably argue" for Heine and Kuteva's construal of CIG as an RL-agentive direct replication of a grammaticalization process because of his conviction that the phenomenon is "largely driven by effort reducing practices of which speakers are only marginally aware" (**Ross2007**: 135).

Matras (**Matras2009**; **Matras2011gram**), however, supports Heine and Kuteva's initial characterization by arguing for RL-agentivity in his own recent accounts of CIG, which provide more attention to the role played by the individual bilingual in the phenomenon's actuation. He cites the individual's communicative imperatives and creative impulse as the primary force driving the replication of grammaticalization processes, as speakers actively borrow from constructions they control in one of their languages as a source of expressive innovation in the other, limiting this transfer solely to "pattern" out of respect for the norms of the distinct speech communities in which they operate. Matras' account thus has the benefit of aligning with the motivating forces theorized to obtain for grammaticalization processes more generally. As described by Lehmann in his consideration of grammaticalization's communicative/pragmatic dimension:

> To the degree that language activity is truly creative, it is no exaggeration to say that languages change because speakers want to change them … they do not want to express themselves the same way they did yesterday, and in particular not the way that somebody else did yesterday (**Lehmann1985**).

Building upon this position, it holds that in scenarios of language or dialect contact innovating speakers may very well wish to express themselves the same way somebody else did yesterday if the means of expression involved are novel to a distinct speech community with which they are interacting today. This synergy with less controversial understandings of grammaticalization outside the context of language/dialect contact provides a viable counterpoint to the skepticism voiced by Ross, and strongly recommends the association of CIG with RL-agentivity.

In the case of contact between closely related varieties, this characterization of CIG may be further qualified. Under Matras' RL-agentive formulation, pattern

replication occurs in the presence of constraints against the usage of an SL's forms due to speaker expectation and "language loyalty" among members of the RL community (2011: 283). In the broader language contact literature, such sociolinguistic constraints have been noted to play a role in pattern diffusion (e.g., **Epps2005**), and presumably interact with speakers' judgments of interlocutors' perceived bilingual competency in favoring or disfavoring matter-based mixing or borrowing (cf. **Grosjean2001**). In contexts where mutual comprehensibility is a less salient concern, the drivers of pattern vs. matter-based innovation would be expected to be almost purely sociolinguistic and pragmatic, and are perhaps most fruitfully understood through the lenses of indexicality (**Silverstein2003**) and focus at the level of the speech community (**Lepage1985**) rather than as a desire to adhere to a reified linguistic code. Such is the state of affairs most likely to obtain in the case of CIG between neighboring varieties of Arabic, to which we now turn in detail.

## 2 CIG in the development of Arabic future tense markers

### 2.1 Methods of investigation

In the following subsections, I present evidence for the role of CIG as the primary mechanism underlying the development and distribution of future tense markers in the modern Arabic dialects. The data considered is drawn from a survey of eighty-one geographic sample points spanning the contiguous Arabic-speaking world, based on a total of eighty-eight descriptive sources.[2] This sample was constructed as part of the broader investigation of CIG between Arabic dialects presented by **Leddy-Cecere2018**, which investigates the role of CIG in the development of a number of morphosyntactic features in modern Arabic varieties, including future tense markers, genitive exponents, and temporal adverbs meaning 'now'. A discussion of data and findings for the first of these features is the focus of this chapter, and these shall be seen to argue strongly for the identification of CIG as a key force in shaping the evolution of modern Arabic dialects. Readers are encouraged to refer to **Leddy-Cecere2018** for additional examination and expansion of the points to follow.

To begin, I first examine the complete set of specific grammaticalizations of future tense markers attested by the Arabic dialect data. These have been identified via the observation of concurrent processes of desemanticization, extension,

---

[2]Details of sample composition and sources consulted, as well as the selection criteria for each, may be found in Leddy-Cecere (**Leddy-Cecere2018**: 34–35, 43–46).

decategorialization and erosion (as described in §1.2). These individual instances are further sorted into higher level groupings by grammaticalization path: future tense markers deriving from motion verbs meaning 'go', for example, as against those deriving from purposive constructions, etc. Special attention is paid to those specific grammaticalization paths represented by multiple evolutions involving distinct etyma, thus identifiable as potential candidates for the products of replication through CIG. As a final step in the evaluation, the geographic incidence of forms representing such multiply attested paths is considered to assess whether their modern distribution is consistent with a historical account of diffusion via contact. This latter portion of the analysis will be presented in §2.3, following the complete accounting of grammaticalized forms provided in §2.2 immediately below.[3]

## 2.2 Grammaticalizations of Arabic future tense markers by grammaticalization path

### 2.2.1 Futures from 'go' (FUT < GO)

Grammaticalizations of future tense markers from forms of lexical verbs meaning 'go' are well represented in the Arabic data. This grammaticalization path is widely attested cross-linguistically, providing one of the major sources for the development of future tense markers worldwide (**Bybee1994**; **Heine2002**). Grammaticalizations of specific items observed in the cross-dialectal Arabic sample are described below.

*rāyiḥ

Future markers representing grammaticalized forms of an active participle *rāyiḥ 'going' are found across a broad east-west swath of the Arabic-speaking world, extending from southern Iraq in the east to Algerian territory in the west. Differing degrees of grammaticalization are attested, with some forms maintaining full phonological integrity and categorial membership (e.g. Basra *rāyiḥ* (**Mahdi1985**), which displays adjectival gender/number agreement with its subject) and others showing dramatic erosion and of loss of morphosyntactic autonomy (including Cairo *ḥa-* ~ *ha-*, as described in §1.2). Semantically, some forms, such as Algiers *rāḥ* and Jerusalem *rāyiḥ* ~ *rāḥ* ~ *ḥā-*, are recorded as expressing a value of immediate future or future intent (**Boucherit2011**; **Rosenhouse2011**), while the

---

[3]For further discussion and justification of each stage of this heuristic, see Leddy-Cecere (**Leddy-Cecere2018**: 36–43, 209–214).

majority are associated with a meaning of general futurity.

\*ɣādī

Grammaticalized forms of the active participle \*ɣādī 'going' are common future markers throughout much of Morocco and adjacent regions. Reduced and invariant forms often co-exist alongside less grammaticalized reflexes, thereby attesting discrete links in an increasingly advanced grammaticalization chain, as seen in Casablanca *ɣādī ~ ɣa-* (**caubet2011**).

\*māšī

Grammaticalized future markers deriving from the active participle \*māšī 'going' occur in two distinct geographic pockets, one centered in north-central Morocco and the other in Tunisia and eastern Algeria. In addition to more predictable effects of phonetic erosion and decategorialization, several forms from the latter area display a further example of irregular sound change with the sporadic denasalization of \*/m/ > /b/, as in Sousse *māš ~ bāš* (**Talmoudi1980**).

\*sāyir

Alone in the sample, Maltese attests a future marker deriving from a grammaticalized form of the active participle \*sāyir 'going'. This can be found in both an inflecting form *sejjer* (MSG)/*sejra* (FSG)/*sejrin* (PL) and as the more grammaticalized, invariant forms *ser* and *se* (**Vanhove1993**).

### 2.2.2 Futures from 'want' (FUT < WANT)

Grammaticalizations from source constructions indicating desire or volition are another cross-linguistically common origin for future tense markers (**Bybee1994**; **Heine2002**), and are similarly widespread to their FUT < GO counterparts in the Arabic dialect data. Specific grammaticalizations are discussed below.

\*yabɣā ~ yabɣī

Grammaticalized forms of the imperfective verb \*yabɣā ~ yabɣī 'want' serve as future markers across a large portion of the Arabic-speaking world, stretching from the Arabian Peninsula across the Red Sea to the greater Sudanic area and then northward through modern Libya. While many Arabic varieties attest only a highly grammaticalized, reduced form of the item

(e.g., Abu Dhabi *b-*; **Qafisheh1977**), other dialects display direct evidence of multiple stages of the grammaticalization chain, e.g. Ḥarb *yabɣā ~ yabā ~ ba-* (**Il-Hazmy1975**).

*biddu ~ widdu

Future tense markers arising from grammaticalizations of *biddu ~ widdu 'want' are found throughout the broader Levantine area. In their most phonetically reduced forms (e.g., Soukhne *b-* (**Behnstedt1994**)), they are often superficially indistinguishable from the highly grammaticalized products of *yabɣā ~ yabɣī discussed above; several dialects, however, provide clear evidence for a distinct chain of development, such as Jebel Ansariye *baddo ~ bado ~ b-* (**Lewin1969**) and Cilicia *baddu ~ baddi- ~ bad-* (**Procházka2011**). In some varieties, grammaticalizations of *biddu ~ widdu operate alongside other markers of future tense to designate a more specified value: Damascus *bəddo ~ b-* is often reported to denote a modal value of possible or planned future, as opposed to the *rāyiḥ-derived forms *raḥ ~ ḥa-* which indicate a higher degree of certainty or expectation (cf. **Lentin2011**). In other dialects, these forms would appear to have further desemanticized and extended to a value of more general futurity. Future investigation is needed into the degree to which reduced reflexes of *biddu may have merged in mental representation with the continuous aspect marker *b-* present in many of the same varieties; relevant parallels might be drawn with scenarios of near homophony like that found in the dialect of Dhofar, in which continuous *bi-* exists alongside future *bā-* (<*yabɣā ~ yabɣī) (**Davey2016**).

*yišā

In a number of Yemeni dialects, the future tense marker may be traced to a grammaticalized form of *yišā 'want'. It is notable that in cases such as Sana'a *ša-* this form is used only with the first person singular verb (**Watson1993**); in such circumstances, it is possible that its ultimate source should be more properly identified with *ašā 'I want'.

*ydawr

Varieties belonging to the Ḥassāniyya dialect complex of Mauritania and neighboring Mali are recorded as utilizing a grammaticalized form of the verb *ydawr 'want' with a following imperfective verb to denote a value of

intentional future. This grammaticalization is relatively "light," consisting primarily of desemanticization and extension with little in the way of de-categorialization or erosion: Nouakchott *ydoːr*, for example, denotes future intent while continuing to operate morphosyntactically as a fully inflected finite verb (**Taine-Cheikh2011**).

*bɣā

Dialects of southern Morocco and southwestern Algeria occasionally attest grammaticalized forms of *bɣā 'want' expressing a future tense value. Though lexically similar in origin to the grammaticalizations based on *yabɣā ~ yibbā ~ yibbī discussed above, the phonological shape of these items (e.g. Marrakech *bɣaː ~ ba-* (**Sánchez2014**)) recommends an identification of their source in the perfective stem *bɣā, which is the typical means for expressing 'want' in this area.

### 2.2.3   Futures from 'come' (FUT < COME)

Another cross-linguistically common path of future tense grammaticalization, that involving verbs meaning 'come, return' (**Bybee1994**; **Heine2002**), is represented in the Arabic data by markers originating from a single source etymon, *ʕād 'return'.

*ʕād

Future tense markers traditionally identified as grammaticalized forms of *ʕād 'return' are attested in three locations in the cross-dialectal survey: Yemen, Upper Egypt and interior Tunisia. The forms found in Tunisia and Egypt, Tozeur *ʕa-* and Aswan *ʕa-* (**Saada1984**; **Schroepfer2019**), are highly reduced, and thus difficult to ascribe definitively to a specific source. It is notable that in both of these dialects the markers in question vary with a 'go'-derived future *ḥa-* and could thus plausibly represent an erosion of the latter in the form of a sporadic lenition of /ḥ/ > /ʕ/ (not to mention that Aswan *ʕa-* on its own might be linked to local *ʕāyiz ~ ʕāwiz* 'want'). At least in the case of the Yemeni forms, however, an origin in *ʕād seems clear, as reduced forms such as Sana'a *ʕā-* display an allomorph *ʕad-* in prevocalic contexts (**Watson1993**).

### 2.2.4   Futures from purposive constructions (FUT < PURP)

A further source of future tense markers in the Arabic data involves the grammaticalization of purposive operators. This path is not widely discussed in the

cross-linguistic grammaticalization literature, though intriguingly the reverse trajectory, that of PURP < FUT, is noted (**Bybee1994**). The primary difficulty would seem to rest in the identification of a clear process of desemanticization, as it is difficult to judge precisely which function between FUT and PURP is more concrete/abstract than the other. Despite this, the occurrence of extension, decategorialization and erosion in the Arabic forms seems to recommend their identification as products of a grammaticalization process.

*ḥattā

> Grammaticalizations of *ḥattā 'in order to' are used to indicate future tense in areas of northern Mesopotamia, the coastal Levant, and Oman. In terms of geographic distribution and the specific path of phonetic erosion followed, it may be possible to recognize Levantine and Mesopotamian forms like Cypriot Maronite *tta-* and Mosul *də-* (**Borg1985**; **Jastrow1979**) as representing a single historical innovation, though Oman *ḥa- ~ ha-* is more likely an independent development. In the Omani case, the attested use of *ḥa-* with purposive meaning recommends a source in *ḥattā rather than GO-future *rāyiḥ: *šrab ḥa-turwe!* 'Drink so your thirst be quenched!' (**Reinhardt1894**).

### 2.2.5 Futures from 'to busy oneself with' (FUT < VERB OF ACTIVITY/PREPARATION)

A small number of Arabic dialects utilize a future tense marker seeming to derive from a grammaticalized form of a verb meaning 'to busy oneself'. Such a path of development is not discussed in the cross-linguistic literature on grammaticalization, but perhaps has a counterpart in the use of grammaticalized Southern American English *fixing to ~ fixin' a ~ fi'na* to express proximate futurity (cf. **Wolfram1998**). In any case, obvious desemanticization, extension, decategorialization and erosion of the source form indicate a clear example of grammaticalization in this case.

*lāhī

> In the Ḥassāniyya dialects of Mauritania, Mali and far southern Morocco, the future tense marker derives from a grammaticalized form of *lāhī, itself the active participle form of the verb *lha* 'to busy oneself'. Decategorialization is attested in all cases by the lack of adjectival agreement marking predicted for the original participial, and in at least some varieties phonetic erosion is evidenced as well: Mali *lāhi ~ lā* (**Heath2003**).

## 2.3 Evidence of replication and diffusion via contact

Of the five grammaticalization paths for Arabic future markers presented above, two merit closer examination in the search for evidence of CIG: those of FUT < GO and FUT < WANT. These paths are identified due to the fact that each is represented in the data by multiple, parallel realizations arising from etymologically distinct but semantically and functionally analogous sources. Such a state of affairs plausibly reflects the result of continued processes of replication, whereby a grammaticalization process occurring in one Arabic variety is transferred to another and recreated using native etymological material.

Both paths identified, however – together representing the great majority of future tense markers attested in the sample – are also extremely common cross-linguistically, and could conceivably have fed multiple independent developments instantiated across the modern Arabic dialect continuum. Key to selecting between an analysis of CIG and one of repeated, internally-motivated grammaticalization is the factor of geography, as in the absence of fine-grained historical sociolinguistic data (see §1.3) this is perhaps the most reliable proxy in positing the feasibility of historical contact between dialects. In the case of CIG, analogous grammaticalization processes ought to be positioned in a geographically contiguous (or near contiguous) bloc, consistent with a history of diffusion via contact between speakers of neighboring dialects. In a scenario of independent development, on the other hand, one should expect the various grammaticalizations to be more or less randomly distributed across the map, equally likely to occur in any individual dialect considered.

The geographic incidences of the members of the FUT < GO and FUT < WANT paths both clearly align with the contiguous profile anticipated for the results of CIG. All realizations of FUT < GO future tense markers described connect geographically with other members of the bloc. The large eastern and central zone of *rāyiḥ futures, encompassing southern Mesopotamia, much of the Levant, and the Nile Valley, stretches westward across Libya (where *rāyiḥ-derived forms are recorded alongside *yibbī-based WANT-futures) to include most of Algeria. Directly adjacent to this North African arm of the *rāyiḥ forms are found grammaticalizations of *ɣādī in Morocco and of *māšī in Tunisia. Further neighboring or co-territorial with the latter two areas are a second set of *māšī-based forms in northern Morocco and the Maltese *sāyir-derived future tense marker, thus completing the connected geographic trend. Future markers representing the path FUT < WANT display a similar spatial contiguity. Grammaticalizations of *biddu ~ widdu in the Levant stretch to reach those of *yabɣā ~ yabɣī present in the Arabian Peninsula. These in turn span the Red Sea across to the greater Sudanic

area and northward through the central Sahara into Libya. Moving to the west and southwest of this zone, the next future markers encountered include grammaticalizations of *byā and *ydawr, respectively. Rounding out the set, forms derived from *yišā exist in close proximity to analogous *yabɣā ~ yabɣī futures in Yemeni territory. While the integrity of this WANT-future bloc may seem to be challenged by natural features such as the Red Sea and the Sahara Desert, historical and anthropological investigations of the regions in question have rather shown persistent social and cultural connectivity across these would-be barriers (**Power2012**; **Lydon2009**). This evaluation is supported by the distribution of additional Arabic dialectological isoglosses extending beyond the discussion of CIG.

The geographic contiguity displayed by the representatives of both the FUT < GO and FUT < WANT pathways favors an interpretation of areal diffusion over one of independent, internally motivated occurrence (of the type perhaps evidenced by the more scattered distributions of the sole representatives of FUT < COME and FUT < PURP). The optimal account for the development of the modern Arabic GO- and WANT-futures, together representing the greater part of future tense markers attested in the data, is thus one by which grammaticalization processes leading to the development of new future tense markers have repeatedly been subject to transfer and replication between speakers of neighboring dialects. A CIG-driven analysis such as this has the benefit of accounting for both the development of individual dialect forms and more global trends in source semantics and geographic incidence, and offers a theoretically unified interpretation of the Arabic data obtaining on multiple scales.

## 3 Conclusion

The analysis summarized above has demonstrated the significant explanatory power of CIG as an account for the development of Arabic future tense markers. Additional proposed occurrences of CIG between Arabic dialects, pertaining to genitive exponents and temporal adverbs meaning 'now', are identified and examined in **Leddy-Cecere2018**. Together with the future tense data discussed here, these call for corroboration and refinement at the hands of future investigators.

Should further examination provide evidence for a widespread history of CIG between Arabic dialects, this finding could prove instrumental in satisfactorily accounting for a number of so-called "pluriform" developments which have repeatedly vexed students of Arabic dialectology. Defined by Versteegh as functionally analogous but etymologically disparate developments for which "a general

trend … has occurred in all Arabic dialects, and an individual translation of this trend in each area," dialect contact has most often been dismissed as a causal mechanism for these innovations due to a belief that "typically dialect contact leads to the borrowing of another dialect's markers, not to the borrowing of a structure that is then filled independently" (**Versteegh2001**: 108). CIG provides a theoretical mechanism by which precisely such borrowing and filling may occur, and as such offers the dialectologist a novel analytical tool in the elucidation of structural transfer and diffusion between Arabic varieties.

A critical open question in the application of CIG to the Arabic context, as well as in study of the phenomenon more generally, lies in the problem of agentivity and actuation (as discussed in §1.3). Here, too, further accrual of Arabic data has the potential to inform broader domains of inquiry. If Arabic is established as a productive ground for the study of CIG and significant cases of the historical transfer of grammaticalization pathways between dialects are brought to light, it stands to reason that the same societal and linguistic forces which have motivated these to take place may still be in force, and that observation of synchronic Arabic dialect interaction represents a singular opportunity to catch newly occurring instances of CIG "in the act" and to observe their progress in real time (for at least one such attempt already presented, see **Abuamsha2016**). Studies of this type will enable linguists to add critically lacking synchronic data to their sociolinguistic and psycholinguistic analyses of CIG, and so elaborate and strengthen ongoing theorizations of a revelatory new dimension of contact-induced language change.

## Further reading

For a complete theoretical discussion of contact-induced grammaticalization, see Heine and Kuteva (**Heine2003**; **Heine2005**). The former work is an article-length sketch of the proposal and is valuable as a direct and concise reference, while the latter provides a more elaborated description with additional linguistic examples. **Matras2009** provides valuable commentary and critique of Heine and Kuteva's work while simultaneously extending exploration to the psycholinguistic and sociolinguistic dimensions of CIG.

For an overview of grammaticalization processes in the development of Arabic future tense markers (though without reference to contact) see **Stewart1998**. A more detailed treatment of CIG processes in the development of the Arabic future markers and other morphosyntactic features may be found in **Leddy-Cecere2018**.

## Acknowledgements

## Abbreviations

| | |
|---|---|
| CIG | contact-induced grammaticalization |
| FSG | feminine singular |
| FUT | future |
| invar. | invariant |
| MSG | masculine singular |
| PL | plural |
| PURP | purposive |
| RL | recipient language |
| SL | source language |

# Chapter 7

# Contact and the expression of negation

Chris Lucas

??

This chapter presents an overview of developments in the expression of negation in Arabic and a number of its contact languages, focusing on clausal negation, with some remarks also on indefinites in the scope of negation. For most of the developments discussed in this chapter, it is not possible to say for certain that they are contact-induced. But evidence is presented which, cumulatively, points to widespread contact-induced change in this domain being the most plausible interpretation of the data.

## 1 Overview of concepts and terminology

### 1.1 Jespersen's cycle

In the past few decades there has been steadily intensifying interest among linguists in historical developments in the expression of negation, with particular attention given to the fact that these developments typically give the appearance of being cyclical in nature. We can date the beginning of this sustained interest to **Dahl1979**'s (**Dahl1979**) typological survey of negation patterns in the world's languages, in which he coined the term Jespersen's cycle[1] for what is by now the best known set of developments in this domain: the replacement of an original negative morpheme with a newly grammaticalized alternative, after a period in which the two may co-occur, prototypically resulting in a word order shift from preverbal to postverbal negation. The best known examples of Jespersen's

---

[1]The name was chosen in recognition of the early identification of this phenomenon by the Danish linguist Otto Jespersen in a (**??**) article, though others did identify the same set of changes earlier: **Meillet1912**, for example, but also, significantly for the present work, **Gardiner1904**, who observed a parallel set of changes in Coptic and Arabic as well as French (cf. van der **Auwera2009**).

cycle (both supplied, among others, by Jespersen himself in his 1917 work) come
from the history of English (1) and French (2).

(1)   English (**Jespersen1917**)
      Stage I – Old English
      ic ne    secge
      I   NEG say.PRS.1SG
      'I do not say.'

(2)   Stage II – Middle English
      I ne    seye       not
      I NEG say.PRS.1SG NEG
      'I do not say.'

(3)   Stage III – Early Modern English
      I say not

(4)   French (**Jespersen1917**)
      Stage I – Old French
      jeo ne    di
      I    NEG say.PRS.1SG
      'I do not say.'

(5)   Stage II – contemporary written French
      Je ne    dis        pas
      I   NEG say.PRS.1SG NEG
      'I do not say.'

(6)   Stage III – contemporary colloquial French
      Je dis          pas
      I   say.PRES.1SG NEG
      'I do not say.'

   Jespersen's cycle has since been the subject of intensive investigation, espe-
cially in the languages of Europe (e.g. **BerniniRamat1992**; **BerniniRamat1996**;
Willis, **LucasBreitbarth2013**), but also beyond (e.g. **Lucas2007**; 2009; 2013; **LucasLash2010**;
Devos & van der **Auwera2013**; van der **AuweraVossen2015**; 2016; 2017), with a
picture emerging of a marked propensity for instances of Jespersen's cycle to be
areally distributed, as we will see below in the discussion of Jespersen's cycle in
Arabic and its contact languages (§2).

While Jespersen's cycle is the best known, best studied, and perhaps crosslinguistically most frequently occurring set of changes in the expression of negation, two other important types of changes must also be mentioned here: Croft's cycle and changes to indefinites in the scope of negation.

## 1.2 Croft's cycle

In a typologically-oriented (**??**) article, Croft reconstructs from synchronic descriptions of a range of languages a recurring set of cyclical changes in the expression of negation. Unlike Jespersen's cycle, in which the commonest sources of new negators are nominal elements expressing minimal quantities, such as 'step' or 'crumb', or generalizing prounouns like '(any)thing', CROFT'S CYCLE (named for Croft by **Kahrel1996**), involves the evolution of new markers of negation developed from negative existential particles. **Croft1991** distinguishes the following three types of languages:

Type A: the verbal negator is also used to negate existential predicates

Type B: there is a special negative existential predicate distinct from the verbal negator

Type C: there is a special negative existential predicate, and this form is also used to negate verbs

For Type A, **Croft1991** cites the example of Syrian Arabic *mā fī* 'there is not' and *mā baʕref* 'I do not know' among others. For Type B he cites (1991: 9), among other examples, the contrast between the Amharic negative existential *yälläm* (affirmative existential *allu*) and regular verbal negation *a(l)...-əm*. For Type C he cites (1991: 11–12) Manam (Oceanic) among other languages, giving the example in (3).

(7) Manam (**Croft1991**: 11–12; **Lichtenberk1983**: 385, 499)
Verbal negation
tágo      u-lóŋo
NEG(EX) 1SG.RLS-hear
'I did not hear.'

(8) Negative existential predicate
anúa-lo    tamóata tágo      [*i-sóaʔi]
village-in person   NEG.EX [3SG.RLS-EX]
'There is no one in the village.'

A number of languages also exhibit variation between two of the types: A ~ B, B ~ C, and C ~ A. This indicates a cyclical development A > B > C > A,

in which a special negative existential predicate arises in a language (A > B), comes to function also as a verbal negator (B > C), and is then felt to be the negator proper, requiring supplementation by a positive existential predicate in existential constructions (C > A).

While Croft's cycle is less common than Jespersen's cycle, and has not been shown to have occurred in its entirety in the recorded history of any language, we mention it here because recent work by Wilmsen (2014: 174–176; 2016), discussed below in §2.1.2, argues for several instances of Croft's cycle in the history of Arabic.

## 1.3 Changes to indefinites in the scope of negation

The final major set of common changes to be dealt with here involve indefinite pronouns and quantifiers in the scope of negation. Here too cyclical patterns are commonplace, and these changes have been labelled "the argument cycle" (**Ladusaw1993**) or "the quantifier cycle" (**Willis2011**). What we find is that certain items, typically quantifiers such as 'all' or 'one' or generic nouns such as 'person' or 'thing', are liable to develop restrictions on the semantic contexts in which they can occur, namely what are referred to as either downward-entailing or non-veridical contexts (see **Giannakidou1998** for details and the distinction between the two). In essence, this means interrogative, conditional, and negative clauses, as well as the complements of comparative and superlative adjectives, but not ordinary affirmative declarative clauses. Items that are restricted to appearing in such contexts, such as English *ever* (consider the ungrammaticality of, e.g., *\*I've ever been to Japan*), are generally termed NEGATIVE POLARITY ITEMS. Often, however, we find negative polarity items whose appearance is restricted to a subset of these contexts, and much the most common restriction is to negative contexts only. Items with this narrower distribution, such as the English degree adverbial phrase *one bit*, are generally termed strong negative polarity items and those with the wider downward-entailing/non-veridical distribution may be termed weak negative polarity items in contrast.

A commonly recurring diachronic tendency of such items is that they become stronger over time. That is, an item goes from having no restrictions, to being a weak negative polarity item, to being a strong negative polarity item, to eventually being itself inherently negative. The best-known instance of this progression comes from French *personne* 'nobody' and *rien* 'nothing'. These derive from the ordinary, unrestricted Latin generic nouns *persona* 'person' and *rem* 'thing' and still behaved as such in medieval French, as in (4).

(9) Medieval French (**Hansen2013**: 72; **Buridant2000**: 610)

Et si vous dirai une rien.
and so 2PL say.FUT.1SG INDEF.SG.F thing

'And so I'll tell you a thing.'

In later medieval French they grammaticalized as indefinite pronouns and began to acquire a weak negative polarity distribution, as in the interrogative example in (5).

(10) C13th French (**Hansen2013**: 72; **Buridant2000**: 610)

As tu rien fet?
AUX.PRES.2SG 2SG.SBJ anything do.PP

'Have you done anything?'

In present-day French these items have become essentially inherently negative, as shown in (6). They can no longer appear in interrogative, conditional or main declarative clauses with an affirmative interpretation (**Hansen2013**), though an affirmative interpretation remains possible in comparative complements, albeit largely in frozen expressions, as in *rien au monde* 'anything in the world' in (7).

(11) Contemporary French (**Hansen2013**)

Qui t' a vu? Personne!
who 2SG.OBJ AUX.PRES.3SG see.PP nobody

'Who saw you? Nobody!'

(12) Contemporary French (**Hansen2013**)

J' aime le vin mieux que rien au monde.
1SG.SBJ DEF wine better than anything in+DEF.SG.M world

'I like wine better than anything in the world'

Note that French *rien* 'nobody' and *personne* 'nothing', like their equivalents in many other Romance varieties (e.g. Italian *niente* and *nessuno*), are not straightforward negative quantifiers like English *nobody* and *nothing*, even disregarding their behaviour in contexts such as (7). This is because French, like many other languages but unlike standard English, standard German, classical Latin etc., exhibits NEGATIVE CONCORD. This refers to the fact that when two (or more) elements which express negation on their own co-occur in a clause the result is not logical double negation (i.e. a positive) but a single logical negative, as illustrated

in (8).

(13)  Contemporary French (**Hansen2013**)
Personne n'   a        rien   dit
nobody  NEG AUX.PRES.3SG nothing say.PP
'Nobody said anything.'

Items which have this unstable behaviour are distinguished from straightforward negative items by the term N-WORD (coined by **Laka1990**; see also **Giannakidou2006**). We will see in §3 that these distinctions and terminology are helpful in understanding developments in varieties of Arabic and its contact languages that directly parallel those described above for French.

## 2 Developments in the expression of clausal negation

### 2.1 Arabic

### 2.1.1 Synchronic description

One of the most striking ways that a number of spoken Arabic varieties differ from Classical and Modern Standard Arabic is in the expression of negation. In Classical and Modern Standard Arabic, and in the majority of varieties spoken outside of North Africa, negation is exclusively preverbal, with the basic verbal negator in the spoken varieties being *mā*, as in the Damascus Arabic example in (9).

(14)  Damascus Arabic (**Cowell1964**)
hayy masʔale mā  bəddaḥḥək
this.F matter  NEG laugh.CAUSE.IMPF.INDIC.3SG.M
'This is not a laughing matter.' (lit. 'does not cause laughter')

But in the varieties spoken across the whole of coastal North Africa and into the southwestern Levant, as well as in parts of the southern Arabian Peninsula (see **Diem2014**; **Lucas2018** for more precise details), negation is bipartite, with preverbal *mā* joined by an enclitic *-š* which follows any direct or indirect pronominal object clitics, as in (10).

(15)  Cairo Arabic (advertising slogan)

banda ma   yit?al-lahā-š                                    la?
Panda NEG say.PASS.IMPF.3SG.M-DAT.3SG.F-NEG no

'You don't say "no" to Panda.' (lit. 'Panda, "no" is not said to it')

Finally, in a subset of the varieties that permit the bipartite construction in (10), a purely postverbal construction is also possible, as in the Palestinian Arabic example in (11).

(16)   Palestinian Arabic (**Seeger2013**)
       badaḫḫin[i]-š
       smoke.IMPF.INDIC.1SG-NEG

       'I don't smoke.'

### 2.1.2 Jespersen or Croft?

There is near unanimous agreement among those who have considered the matter that the bipartite construction illustrated in (10) arose from the preverbal construction via grammaticalization, phonetic reduction, and cliticization of *šay?* 'thing', and that the purely postverbal construction in (11) in turn arose from the bipartite construction via omission of the original negator *mā*. As such, Lucas (2007; 2009; 2018) and **Diem2014**, among many others, view this as a paradigmatic case of Jespersen's cycle.

The only dissenting voice is that of Wilmsen (2013; 2014), who describes the parallels between the Arabic data and that of well known cases of Jespersen's cycle such as French as being "dutifully mentioned by all" (2014: 117) who write on the topic. **Wilmsen2014** turns the agreed etymology of negative *-š* on its head by arguing: (i) that the original form in Arabic was *šī*, not *šay?*;[2] (ii) that at an early stage this form had the full range of functions that we observe for it in different Arabic dialects today (existential predicate, indefinite determiner, interrogative particle; see **Wilmsen2014**: ch.3, 122–123); (iii) that this element was then reanalysed as a negative particle; and (iv) *šī/šay?* as a content word 'thing' is a later development of the function word – an instance of degrammaticalization. For a discussion of some of the numerous difficulties with these proposals, see Al-**Jallad2015**, Pat-**El2016**, **Souag2016**, and **Lucas2018**.

---

[2]**Wilmsen2014** also attempts to trace his etymology back further to the Proto-Semitic third-person pronouns. Apart from the implausability of the putative semantic shift from definite pronoun to indefinite determiner, this reconstruction is untenable on phonological grounds (see Al-**Jallad2015** for details).

A specific element of Wilmsen's proposals that merits our attention here is his suggestion that, while in his view we should not see the developments in Arabic as an instance of Jespersen's cycle, we can discern in them an instance of Croft's cycle. As we will see below, this suggestion involves a distortion or misunderstanding of both the Arabic data and the sorts of patterns that constitute genuine instances of Croft's cycle, but the proposal has some *prima facie* plausability, because of the existence in some dialects of the south and east of the Arabian Peninsula of an existential predicate *šī/šē/šay*, as in (12).

(17)   Northern Omani Arabic (**Eades2009**)
hmīr      šē... l-ḥmīr      barra
donkey.PL EX   DEF-donkey.PL outside
'There were donkeys... the donkeys were outside.'

Note that a similar element *śī* /ɬiː/, with the same existential function, is found in the Modern South Arabian languages (MSAL) of Yemen and Oman, as in (13) from Mehri of Yemen.

(18)   Mehri of Yemen (**Watson2011**)
śī   fśē
EX lunch
'Is there any lunch?'

Though **Wilmsen2014** seems to view Arabic *šī* and Modern South Arabian *śī* as cognates, it is more likely that the presence of this item in the one set of varieties is the result of transfer from the other (cf. Al-**Jallad2015**). The direction of transfer is unclear, however. At first glance, the fact that *śī* as an affirmative existential is found in essentially all of the MSAL spoken on the Arabian Peninsula, which have a long history of intensive contact with Arabic, but not in Soqotri, spoken on the island of Soqotra, where contact with Arabic is more recent and less intensive (Simeone-**Senelle2003**), would appear to suggest that this is an innovation within Arabic originally, which was then transferred to just those MSAL with which there was most contact. On the other hand, the precise situation in Soqotri is perhaps instructive. Here the affirmative existential predicate is a unique form *ino*, while the negative existential predicate is *biśi (Simeone-Senelle2011: 1108)*. It is conceivable that the latter is a borrowing from Arabic, since affirmative existentials in *b-* are widespread in the Arabic dialects of Yemen. But a negative existential predicate *bīši* or similar is completely unattested in the Yemeni data provided by Behnstedt (2016: 346–348). This suggests,

therefore, that: (i) existential *śī* is an original feature of MSAL; (ii) Soqotri is an example of a Type B language in Croft's typology, having innovated a new affirmative existential predicate *ino*, such that there is a special negative existential predicate that is neither identical to the verbal negator, nor simply a combination of the verbal negator with the affirmative existential predicate; and (iii) *šī* as an existential predicate in Arabic dialects is the result of transfer of MSAL *śī*.

This scenario is supported by the distribution of existential *šī* within Arabic varieties: the only clear cases are in dialects of Yemen and Oman with a history of contact with MSAL, and dialects of the Gulf whose speakers are known to have migrated there from Yemen or Oman (such as Šiḥḥī, see §2.4 below). In various places Wilmsen tries to make a case for existential uses of *šī* outside this region, but this appears to be the result of confusion on his part between *šī* as a *bona fide* existential predicate and the existential presupposition that will inevitably be associated with the use of *šī* as an indefinite determiner (see e.g. **Heim1988** on the semantics of indefinite noun phrases). For example, **Wilmsen2014** cites Caubet's (1993a: 123; 1993b: 280) Moroccan Arabic examples in (14) as evidence of an existential use of *šī* as far west as Morocco. But there is no justification for Wilmsen's contradicting Caubet's uncontroversial analysis of *šī* as an indefinite determiner here: there are no existential predicates in these examples – the existence of the referents of the indefinite noun phrases is presupposed, not asserted.

(19)  Moroccan Arabic (**Caubet1993a**: 123; 1993b: 280)
      ši      nās                kayāklu-ha
      INDEF people EAT.IMPF.3PL-3FSG
      'Some people eat it.'

(20)  ši      nās                kaybɣēw əl-lbən
      INDEF people LIKE.IMPF.3PL DEF-milk
      'Some people like milk.'

Nevertheless, *šī* does function as an existential predicate in a few Arabic varieties. The question, then, is whether a negated form of this predicate participates in a version of Croft's cycle, as Wilmsen maintains.

For the vast majority of Arabic varieties the answer is a clear no: these varieties straightforwardly belong to Type A of Croft's typology. The verbal negator (*mā*, *mā…-š*, or *-š*) is also used to negate existential predicates, as illustrated in (15) for Cairo Arabic.

(21)  Cairo Arabic

ma ʕamalt[i]-š ḥāga
NEG do.PRF.1SG-NEG thing
'I didn't do anything.'

(22) ma fī-š ḥāga
NEG EX-NEG thing
'There is nothing.'

Wilmsen (2014: 173–175) suggests that Type B and Type C constructions can also be found, however. For Type B ("there is a special negative existential predicate, distinct from the verbal negator"; **Croft1991**: 6), he cites Ṣanʕānī *māšī* and Moroccan *māši*. Ṣanʕānī *māšī* is certainly a negative existential predicate. But there is nothing special about it – it is a paradigmatic Type A construction, with the negation of the existential predicate (*šī*) performed by the verbal negator (*mā*). Moroccan *māši*, on the other hand, is the negator for nominal predicates (equivalent to *muš/miš/mū* in dialects east of Morocco). It is not a negative existential predicate at all, and, as discussed above, the /ši/ component of this item does not function as an existential in Moroccan, unlike in Ṣanʕānī and other southern Arabian varieties. The existence of *māši* in Moroccan Arabic is thus irrelevant to the question of whether this constitutes a Type B variety.[3] Moroccan is a Type A variety: the positive existential predicate is *kāyn* and it is negated with the ordinary Moroccan verbal negator *ma...-š* (**Caubet2011**). Wilmsen's identification of Arabic varieties of Type C ("there is a special negative existential predicate, which is identical to the verbal negator"; **Croft1991**: 6) depends on the idea that the Arabic predicate negator *māši/muš/miš/mū* is a negative existential predicate, which, as we have seen, it is not. If it were, it would be true that there are Arabic varieties that are optionally of Type C, since in Cairo Arabic, among other varieties, it is possible to negate verbs with *miš* instead of the usual *ma...-š*, as **Mughazy2003** and others have pointed out. But Cairo *miš* (and Moroccan *māši*) are not negative existential predicates, and there is no evidence to suggest they ever were. Moreover, since the Ṣanʕānī negative existential predicate *māšī* also does not seem to be able to function as a verbal negator, there is little apparent merit in **Wilmsen2014**'s (**Wilmsen2014**) attempt to recast the history of negation in Arabic as an instance of Croft's cycle.[4]

---

[3]Van **Gelderen2018** argues that the definition of Croft's cycle should be expanded to encompass cases in which new negators arise from the univerbation of verbal negators with copulas and auxiliaries, as well as existentials. **Wilmsen2014**'s (**Wilmsen2014**) presentation of Croft's Cycle makes no mention of any predicates other than existentials participating in the cycle, however.

[4]This is not to deny, however, that some Arabic dialects show some incipient Type B ten-

### 2.1.3  Internal or external?

It is clear from the above discussion that there is no reason to doubt the majority view of the emergence of negative *-š* as an instance of Jespersen's cycle. What is less clear and more controversial is the question of whether language contact played a role in triggering these developments, or whether this was a purely internal phenomenon (cf. **Diem2014**: 11–12). This is an issue about which it is impossible to be certain given our present state of knowledge. **LucasLash2010** make the case that contact did play a triggering role, however, and also provide arguments against the widely held view that, in the words of **Lass1997**, "an endogenous explanation of a phenomenon is more parsimonious [than one invoking contact – CL], because endogenous change must occur in any case, whereas borrowing is never necessary" (cf. also **Lucas2009**: 38–43). Aside from this generalized reluctance to invoke contact in explanations of linguistic change unless absolutely necessary, another factor that is likely operative in the preference for seeing the Arabic developments as a purely internal phenomenon is ignorance of the wider picture of negative developments in Arabic and its contact languages. It is scarcely an exaggeration to say that everywhere an Arabic variety with bipartite negation is spoken, there is (or was) a contact language that also has bipartite negation, and – just as importantly – wherever Arabic dialects have only a single marker of negation, the local contact languages do too. The picture is similar in Europe and Ethiopia (**Lucas2009**), Vietnam (van der **AuweraVossen2015**), and many other places besides. There can therefore be no doubt that negative constructions, and especially bipartite negation (and hence Jespersen's cycle more generally), are particularly prone to diffusing through languages in contact. In the following sections I will briefly survey apparent instances of transfer of bipartite or postverbal negation in Arabic and Coptic, Arabic and MSAL, Arabic and Kumzari, Arabic and Berber, and Arabic and Domari. For more details see Lucas (2007; 2009; 2013) and **LucasLash2010**.

---

dencies of a different kind. For example, **Behnstedt2016** cites the northern Yemeni dialects of Rās Maḥall as-Sūdeh, Ḥammām ʿAlī and Afk, as varieties in which different morphemes are used in positive and negative existentials, albeit the negative construction used in each case is identical to that used for ordinary verbal negation. In a different context, Stefano Manfredi (personal communication) points out that many urban speakers of Sudanese Arabic use the item *māfīš*, borrowed from Egyptian Arabic, as a negative existential, while ordinary verbal negation is performed with preverbal *mā* alone (without postverbal *-š*).

## 2.2 Arabic and Coptic

Based on an examination of evidence from Judaeo-Arabic documents preserved in the Cairo Genizah, among other sources of evidence, **Diem2014** comes to the conclusion that the Arabic bipartite negative construction found across coastal North Africa originated in Egypt between the 10th and 11th centuries. This chronology and point of origin conforms closely with the conclusions I have drawn on this point in my own work (**Lucas2007**; 2009; **LucasLash2010**), except that I have argued that what triggered the development of bipartite negation in Egypt was contact with Coptic (the name for the Egyptian language from the 1st century CE onwards), which, at the relevant period, had a frequently-occurring bipartite construction *ən…an*, as illustrated in (16).

(23)   Coptic (**LucasLash2010**: 389)
    en   ti-na-tsabo-ou      an   e-amənte
    NEG 1SG-FUT-teach-3PL NEG on-hell
    'I will not teach them about hell.'

The argument made in **LucasLash2010** is that native speakers of Coptic acquiring Arabic as a second language must have encountered sentences negated with preverbal *mā* only, but which also contained after the verb *šī/šāy*, functioning either as an argument '(any)thing' or an adverb 'at all',[5] and interpreted this as the second element of the bipartite negative construction that their first-language Coptic predisposed them to expect. If this is correct, then the initial transfer of bipartite negation from Coptic to Arabic in Egypt should be understood as an instance of imposition under source-language (SL) agentivity, in the terms of Van Coetsem (1988; 2000), while the presence of bipartite negation in the dialects spoken across the rest of coastal North Africa, and the southwestern Levant, should be understood as the result of contact between neighbouring dialects of Arabic.

## 2.3 Arabic and Modern South Arabian

**Diem2014**, like **Obler1990** and, following her, **Lucas2007**, suggests that bipartite negation in the southern Arabian Peninsula must have spread there from Egypt. This is possible, but historical evidence of significant early migration flows in this direction is lacking. The alternative explanation offered by **LucasLash2010**

---

[5]**Diem2014** makes the case that *šī/šāy* had already developed an adverbial use at a very early stage, and that it is this adverbial use that should be seen as the form that was renalaysed as a negator.

is that bipartite negation in the Arabic dialects of this region is an independent parallel development, here triggered by contact with MSAL, all mainland varieties of which have a bipartite negative construction of their own (or once had – some, such as Ḥarsūsi, have largely progressed to stage III of Jespersen's cycle and lost the original preverbal negator), as illustrated in (17) for Omani Mehri.

(24)  Mehri of Oman (**Johnstone1987**)
       əl    təhɛləz          b-ɛy              laʔ
       NEG nag.IMPF.2SG.M with-OBL.1SG NEG

       'Don't nag me!'

If this is correct, then here too, exactly as with the Coptic–Arabic contact in the previous section, we must have had an instance of transfer under SL-agentivity, with MSAL-dominant acquirers of Arabic imposing a bipartite construction on their second-language Arabic by reanalysing *šī/šay* as a negator. The key point is that in all dialects in which *šī/šay* functioned as an indefinite pronoun or adverb 'at all', the potential was there for reanalysis as the second element in a bipartite negative construction. But aside from in the dialects of Egypt and the southern Arabian Peninsula (and latterly dialects adjacent to Egyptian) this reanalysis never took place. Why the reanalysis did take place in Egypt and the southern Peninsula can be understood as being the result of the catalysing effect of contact with languages which themselves had a bipartite negative construction.[6]

## 2.4 Arabic and Kumzari

Kumzari is an Iranian language with heavy influence from both Arabic and MSAL that has only recently been described in detail (see van der Wal Anonby forthcoming). It is spoken on the Musandam Peninsula of northern Oman, where its primary contact language of recent times has been the Šiḥḥī variety of Arabic (see **Bernabela2011** for a sketch grammar), which is clearly of the originally southern Arabian type described by Holes (2016: 18–32).

Šiḥḥī Arabic has no Jespersen stage-II (bipartite) negative construction, but it has both a typical eastern Arabic stage-I construction with *mā*, as in (18)a, perhaps due to recent influence from other Gulf Arabic varieties, alongside a unique

---

[6]For further discussion of the details of these changes, including the issues of the semantics and positioning in the clause of the second negative element in each of the three languages, see Lucas & Lash (2010: 395–401).

(for Arabic) stage-III postverbal construction with *-lu*, as in (18)b. The latter construction is apparently a straightforward transfer of the postverbal negator *laʔ/lɔʔ* of MSAL (cf. (17)).

(25) Šiḥḥī Arabic (**Bernabela2011**)
mā mšēt      ḥaṣāb əl-yōm
NEG go.PRF.1SG Khasab DEF-day
'I didn't go to Khasab today.'

(26) yqōl-lu          bass il-kilmatēn
say.IMPF.3SG.M-NEG only DEF-words.DU
'He doesn't just say the two words.'

The Kumzari negator is the typical Iranian (and Indo-Iranian) *na*. What is less typical is that *na* occurs postverbally in Kumzari, as shown in (19).

(27) Kumzari (van der Wal Anonby forthcoming: 211)
mām-ō      kōr  bur          na
mother-DEF blind become.3SG.RLS NEG
'The mother didn't become blind.'

It seems very likely that contact with Šiḥḥī Arabic has played a role in this shift to postverbal negation, though not enough is known about the historical sociolinguistics of these two speech communities to say with confidence which of the two languages the agents of this change were dominant in.

## 2.5 Arabic and Berber

Berber languages are spoken from the oasis of Siwa in western Egypt in the east, across to Morocco and as far south as Burkina Faso. The most southerly of the Berber varieties – Tashelhiyt, spoken in southern Morocco, Zenaga, spoken in Mauritania, and Tuareg, spoken in southern Algeria and Libya, Niger, Mali and Burkina Faso – have only preverbal negation, as illustrated by the Tuareg example in (20).

(28) Tuareg (**Chaker1996**)
ur   igle
NEG leave.PFV.3SG.M
'He didn't leave.'

These languages have, until recently, either had little significant contact with Arabic, or otherwise only with varieties such as Ḥassāniyya that have only pre-verbal negation with *mā*. All other Berber varieties which are in contact with Arabic varieties with bipartite negation also themselves have bipartite negation, illustrated for Kabyle (Algeria) in (21), or, in a few cases, purely postverbal nega-tion, as in Awjila (Libya), illustrated in (22). The one exception is Siwi, which negates with preverbal *lā* alone – clearly a borrowing from a variety of Arabic, though which variety is not clear (see **Souag2009** for further discussion).

(29)  Kabyle (**Rabhi1996**)
      ul  ittaggad       kra
      NEG fear.AOR.3SG.M NEG
      'He is not afraid.'

(30)  Awjila (**Paradisi1961**)
      akellim iššen-ka          amakan
      servant know.PFV.3SG.M-NEG
      'The servant didn't know the place.'

(31)  Siwi (**Souag2009**)
      lā  gā-nūsd-ak
      NEG FUT-come.1P-DAT.2SG
      'We won't come to you.'

Different Berber varieties have postverbal negators with a range of different forms, but in most cases they either derive from two apparently distinct Proto-Berber items *\*kʸăra* and *\*(h)ară(t)* both meaning 'thing' (**Kossmann2013**), or are transparent loans of Arabic *šay/ši*. This fact, when combined with the respective geographical distributions of single preverbal and bipartite negation in Arabic and Berber varieties, is sufficient to conclude that the presence of bipartite nega-tion in Berber is in large part a result of calquing the second element of the Arabic construction, pace **Brugnatelli1987** and **Lafkioui2013a** (see also **Kossmann2013**: 334; and see **Lucas2007**; 2009 for more detailed discussion).[7] Given that, until recently, native speakers of Arabic in the Maghreb acquiring Berber as a sec-ond language will always have been greatly outnumbered by native speakers of

---

[7]Another postverbal negator – Kabyle *ani* – derives from the word for 'where' (**Rabhi1992**), and so should perhaps be seen as more of an internal development, or at least less directly contact-induced. Tarifit also has a postverbal negator *bu*, whose etymology is uncertain, but which has also been transferred to the Moroccan Arabic dialect of Oujda (**Lafkioui2013b**).

Berber learning Arabic as a second language, we must assume that the agents of this change were Berber-dominant speakers who made the change under RL-agentivity in a process akin to polysemy copying and contact-induced grammaticalization, as identified by **HeineKuteva2005** (see also, Leddy-Cecere, this volume; Manfredi, this volume; Souag, this volume).

## 2.6 Arabic and Domari

The final instance of contact-induced changes to predicate negation to be mentioned here concerns the Jerusalem variety of the Indo-Aryan language Domari, as described by Matras (1999; 2007; 2012; this volume).

Matras (2012: 350–351) describes two syntactic contexts in which negators borrowed from Palestinian Arabic are the only options in this variety of Domari. The first is with Arabic-derived modal auxiliaries that take Arabic suffix inflection, as in *bidd-* 'want' in (24). Here negation is typically with the Palestinian Arabic stage-III construction *-š* (without *mā*), as it is would be also in Palestinian Arabic.

(32)  Jerusalem Domari (**Matras2012**)
      ben-om    bidd-hā-š        žawwiz-hōš-ar
      sister-1SG want-3SG.F-NEG marry.VITR.SUBJ-3SG
      'My sister doesn't want to marry.'

The second is when the negated predicate is nominal, as in (25)a, or, to judge from Matras's examples, when we have narrow focus of negation with ellipsis, as in (25)b. Here the negator that would be used in these contexts in Arabic – *miš* – is transferred to Domari and functions in the same way.

(33)  Jerusalem Domari (**Matras2012**)
      bay-os      mišš kury-a-m-ēk
      mother.3SG NEG  house-OBL.F-LOC-PRED.SG
      'His wife is not at home.'

(34)  day-om      min ʕammān-a-ki       mišš
      mother-1SG from Amman-OBL.F-ABL NEG
      min                      ʕēl-oman-ki day-om
      from FAMILY-1PL-ABL mother-1SG
      'My mother is from Amman, she's not from our family, my mother.'

In addition to these straightforward borrowings, Domari has a bipartite negative construction in which both elements involve inherited lexical material, as illustrated in (26).

(35)  Jerusalem Domari (**Matras2012**)
      ʕašān    ihne ama n-mang-am-san-eʔ      l-ʕarab
      because thus 1SG   NEG-want-1SG-3PL-NEG DEF-Arabs
      'Because of this I don't like the Arabs.'

In Lucas (2013: 413–414) I pointed out that the second element of this construction – *-eʔ* – was apparently not attested in varieties of Domari spoken outside of Palestine, and suggested that its presence in Jerusalem Domari could therefore be the result of influence from the Palestinian bipartite negative construction. Herin (2016; 2018), however, has since convincingly shown that this is incorrect and that the Jerusalem Domari bipartite construction is an internal development with cognates in more northerly varieties, the latter being in contact with Arabic varieties that lack the bipartite negative construction. What is unique about the Jerusalem variety of Domari is that here a stage-III construction with *-eʔ* alone is possible, omitting the original preverbal negator *n(a)* that appears in (26). **Herin2018** argues that it is this stage-III construction, not the stage-II bipartite construction, that should be seen as the result of contact with Palestinian Arabic.

Overall, therefore, while the details naturally vary from one contact scenario to another, we see that negative constructions appear just as liable to be transferred between varieties of Arabic and neighbouring languages as they are between the languages of Europe and beyond.

# 3 Developments in indefinites in the scope of negation

## 3.1 Loaned indefinites

The organization and behaviour of indefinites in the scope of negation seem to be much more resistant to transfer between languages than is the expression of clausal negation, at least in the case of Arabic and its contact languages.[8] Direct borrowing of individual indefinite items is rather common, however. I

---

[8]Though for recent discussion of a related case – namely the acquisition of a determiner function by the Berber indefinite *kra* 'something, anything' via a calque of the polyfunctionality of Maghrebi Arabic *ši* – see **Souag2018**.

make no attempt at an exhaustive list here, but note the following two examples for illustrative purposes.

First, Berber varieties stand out as frequent borrowers of Maghrebi Arabic indefinites. The negative polarity item *ḥadd/ḥədd* 'anyone' is borrowed by at least Siwi (**Souag2009**), Kabyle, Shawiya, Mozabite (**Rabhi1996**), and Tashelhiyt (**Boumalk1996**). The n-word *walu* 'nothing' is borrowed by at least Tarifit (**Lafkioui1996**), Tashelhiyt, and Central Atlas Tamazight (**Boumalk1996**). *ḥətta*, in its function as an n-word determiner is borrowed by at least Tashelhiyt (**Boumalk1996**). *qāʕ*, in its function as a negative polarity adverb 'at all', is borrowed by at least Tarifit and Central Atlas Tamazight (**Boumalk1996**). And the negative polarity adverb *ʕumr* '(n)ever' (< 'age, lifetime') is borrowed by at least Kabyle, Mozabite (**Rabhi1996**), and Tarifit (**Lafkioui1996**). Why these items should have been so freely borrowed, when each of them, with the possible exception of *ḥətta*, have direct native equivalents, is unclear. But it is perhaps to be connected with the high degree of expressivity typically associated with negative statements containing indefinites, which therefore creates a constant need for new and "extravagant" (in the sense of **Haspelmath2000**) means of expressing these meanings.

Second, while Arabic itself seems to have been much more constrained in its borrowing of indefinites from other languages, we can here point at least to the n-word *hīč* 'nothing', which is borrowed from Persian and occurs in both Baghdadi Arabic (**McCarthyRaffouli1964**: 468) and Gulf Arabic (**Holes2001**).

## 3.2 The indefinite system of Maltese

While most or perhaps all Arabic varieties have at least some items that qualify as n-words according to the definition in §1.3, it is only Maltese that has developed into a straightforward negative-concord language with a full series of n-word indefinites in largely complementary distribution with a separate series of indefinites that cannot appear in the scope of negation, as is the situation in French described in §1.3. These two series are shown in Table 1, adapted from Haspelmath & **Caruana1996**.

All the lexical material that makes up the Maltese indefinite system illustrated in Table 1 is inherited from Arabic, but the neat paradigm of n-words for determiner, 'thing', 'person', 'time', and 'place' is much more reminiscent of European Romance languages than Arabic. The extent to which, for example, *xejn* 'nothing' (deriving from *šayʔ* 'thing', with apparent retention of final nunation) is felt by Maltese speakers to be inherently negative,[9] is shown by the existence of the

---

[9]This is despite the fact that it may also occur in interrogatives with non-negative meaning

Table 1: Maltese indefinites

|            | n-words | non-n-words |
|------------|---------|-------------|
| Determiner | ebda    | xi          |
| Thing      | xejn    | xi ħaġa     |
| Person     | ħadd    | xi ħadd     |
| Time       | qatt    | xi darba    |
| Place      | imkien  | xi mkien    |

denominal verb *xejjen* meaning 'to nullify', as illustrated in (27).

(36)  Maltese (**Lucas2013**)
      Iżda xejjen         lilu         nnifs-u
      but  nullify.PRF.3SG.M ACC.3SG.M self-3SG.M
      'But he made himself nothing.'

   As such, it seems likely that the intensive contact that occurred over several centuries between Maltese and the negative-concord languages Sicilian and Italian (cf. Lucas and Čéplö, this volume) played a role in these developments in the Maltese indefinite system. Precisely how this influence was mediated is hard to say, since both borrowing under RL-agentivity and imposition under SL-agentivity were likely operative in the Maltese–Romance contact situation, and either are possible here. See Lucas (**Lucas2013**: 439–444) for further discussion.

## 4 Conclusion

As we have seen, the overall areal picture of bipartite clausal negation in Arabic and its contact languages (and also, to a lesser extent, indefinites in the scope of negation) strongly suggests a series of contact-induced changes, and not a series of purely internally-caused independent parallel developments. What is required in future research on this topic, to the extent that textual and other historical evidence becomes available, is a detailed, case-by-case examination of the linguistic and sociolinguistic conditions under which these constructions emerged in the languages discussed in the course of this chapter.

---

(cf. **CamilleriSadler2017**).

# Further reading

**ChakerCaubet1996** is an edited volume providing a wealth of descriptive data on the expression of negation in a number of Berber and Maghrebi Arabic varieties.

**Diem2014** is a detailed study of the grammaticalization of Arabic *šayʔ* as a negator, with particular attention paid to early sources of textual evidence for this development.

Willis, **LucasBreitbarth2013** is an edited volume containing ten detailed case studies of the history of negation in the languages of Europe and the Mediterranean.

# Acknowledgements

# Abbreviations

1 first person
2 second person
3 third person
ABL ablative
AOR aorist
AUX auxiliary
CAUSE causative
DAT dative
DEF definite
DU dual
EX existential
F feminine
FUT future
IMPF imperfect
INDEF indefinite (article)
INDIC indicative
LOC locative

    MSAL Modern South Arabian language(s)
    NEG negative
    OBJ object
    OBL oblique
    PASS passive
    PL plural
    PP past participle
    PRED predicate
    PRS present
    RL recipient language
    RLS realis
    SBJ subject
    SG singular
    SL source language
    VITR intransitive verb marker

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}

\begin{styleBibliography}
@article{Al-Jallad2015,
author = {Al-Jallad, Ahmad},
journal = {\textit{Bibliotheca} \textit{Orientalis}},
number = {1/2},
pages = {34–46},
title = {What's a caron between friends? {{A}} review article of \citet{Wilmsen201
volume = {72},
year = {2015}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Auwera2009,
address = {Amsterdam},
author = {Auwera, Johan van der},
booktitle = {\textit{Cyclical} \textit{change}},
editor = {Elly Van Gelderen},
```

```
pages = {35–71},
publisher = {Benjamins},
title = {The {Jespersen} Cycles},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{AuweraAuwera2015,
author = {Auwera, Johan van der and Frens Vossen},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Southeast} \textit{
pages = {24–38},
title = {Negatives between Chamic and Bahnaric},
volume = {8},
year = {2015}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{AuweraAuwera2016,
address = {Amsterdam},
author = {Auwera, Johan van der and Frens Vossen},
booktitle = {\textit{Cyclical} \textit{change} \textit{continued}},
editor = {Elly Van Gelderen},
pages = {189–218},
publisher = {Benjamins},
title = {{Jespersen} cycles in {Mayan}, Quechuan and Maipurean languages},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{AuweraAuwera2017,
author = {Auwera, Johan van der and Frens Vossen},
journal = {\textit{Linguistics} \textit{of} \textit{the} \textit{Tibeto-Burman}
```

```
number = {1},
pages = {40–58},
title = {{Kiranti} double negation: {{A}} copula conjecture},
volume = {40},
year = {2017}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Behnstedt2016,
address = {(Trans.) Gwendolin Goldbloom. Leiden},
author = {Behnstedt, Peter},
publisher = {Brill},
title = {\textit{Dialect} \textit{atlas} \textit{of} \textit{{North}} \textit{{Yeme
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@misc{Bernabela2011,
author = {Bernabela, Roy S.},
note = {Leiden University M.A.},
title = {\textit{A} \textit{phonology} \textit{and} \textit{morphology} \textit{sk
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{BerniniBernini1992,
address = {Bologna},
author = {Bernini, Giuliano and Paolo Ramat},
publisher = {Il Mulino},
title = {\textit{{La}} \textit{frase} \textit{negativa} \textit{nelle} \textit{ling
year = {1992}
}
```

\end{styleBibliography}

\begin{styleBibliography}
@book{BerniniBernini1996,
address = {Berlin},
author = {Bernini, Giuliano and Paolo Ramat},
publisher = {Mouton de Gruyter},
title = {\textit{Negative} \textit{sentences} \textit{in} \textit{the} \textit{l
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Boumalk1996,
address = {Paris},
author = {Boumalk, Abdallah},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \texti
editor = {Salem Chaker and Dominique Caubet},
pages = {35–48},
publisher = {L'Harmattan},
title = {{La} négation en berbère marocain},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Brugnatelli1987,
address = {Milan},
author = {Brugnatelli, Vermondo},
booktitle = {\textit{{Atti}} \textit{della} \textit{4a} \textit{giornata} \texti
editor = {Giuliano Bernini and Vermondo Brugnatelli},
pages = {53–62},
publisher = {Unicopli},
title = {{La} negazione discontinua in berbero e in arabo-magrebino},
year = {1987}

```
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Buridant2000,
address = {Paris},
author = {Buridant, Claude},
publisher = {Sedes},
title = {\textit{Grammaire} \textit{nouvelle} \textit{de} \textit{l'ancien} \texti
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{CamilleriCamilleri2017,
address = {Stanford, CA},
author = {Camilleri, Maris and Louisa Sadler},
booktitle = {\textit{Proceedings} \textit{of} \textit{the} \textit{{LFG}'17} \text
editor = {Miriam Butt and Tracy Holloway King},
note = { http://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2
pages = {146–166},
publisher = {CSLI Publications},
title = {Negative sensitive indefinites in Maltese},
year = {2017}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Caubet1993,
address = {1: \textit{Phonologie} \textit{et} \textit{morphosyntaxe}. Paris},
author = {Caubet, Dominique},
publisher = {Peeters},
title = {\textit{L'arabe} \textit{marocain,} vol},
year = {1993a}
}
```

\end{styleBibliography}

\begin{styleBibliography}
@book{Caubet1993,
address = {2: \textit{Syntaxe} \textit{et} \textit{catégories} \textit{grammatic
author = {Caubet, Dominique},
publisher = {Peeters},
title = {\textit{L'arabe} \textit{marocain,} vol},
year = {1993b}
}

\end{styleBibliography}

\begin{styleBibliography}
Caubet, Dominique. 2011. Moroccan Arabic. In Lutz Edzard \& Rudolf de Jong (eds.
\end{styleBibliography}

\begin{styleBibliography}
@incollection{Chaker1996,
address = {Paris},
author = {Chaker, Salem},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \texti
editor = {Salem Chaker and Dominique Caubet},
pages = {9–22},
publisher = {L'Harmattan},
title = {Remarques préliminaires sur la négation en berbère},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{ChakerChaker1996,
address = {Paris},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \texti
editor = {Chaker, Salem and Dominique Caubet},
publisher = {L'Harmattan},

```
title = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{et} \
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Coetsem1988,
address = {Dordrecht},
author = {Coetsem, Frans van},
publisher = {Foris},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two} \
year = {1988}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Coetsem2000,
address = {Heidelberg},
author = {Coetsem, Frans van},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theory}
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Cowell1964,
address = {Washington DC},
author = {Cowell, Mark W.},
publisher = {Georgetown University Press},
title = {\textit{A} \textit{reference} \textit{grammar} \textit{of} \textit{Syrian}
year = {1964}
}

\end{styleBibliography}
```

```
\begin{styleBibliography}
@article{Croft1991,
author = {Croft, William},
journal = {\textit{Journal} \textit{of} \textit{Linguistics}},
number = {1},
pages = {1–27},
title = {The evolution of negation},
volume = {27},
year = {1991}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Dahl1979,
author = {Dahl, Östen},
journal = {\textit{Linguistics}},
pages = {79–106},
sortname = {Dahl, Osten},
title = {Typology of sentence negation},
volume = {17},
year = {1979}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{DevosDevos2013,
author = {Devos, Maud and Johan van der Auwera},
journal = {\textit{Journal} \textit{of} \textit{African} \textit{Languages} \tex
pages = {205–274},
title = {{Jespersen} cycles in {Bantu}: {{D}}ouble and triple negation},
volume = {34},
year = {2013}
}

\end{styleBibliography}
```

```
\begin{styleBibliography}
@book{Diem2014,
address = {Wiesbaden},
author = {Diem, Werner},
publisher = {Harrassowitz},
title = {\textit{Negation} \textit{in} \textit{{Arabic}:} \textit{A} \textit{study}
year = {2014}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Eades2009,
address = {Leiden},
author = {Eades, Domenyk},
booktitle = {\textit{{Arabic}} \textit{dialectology} \textit{in} \textit{honour} \
editor = {Enam Al-Wer and Rudolf De Jong},
pages = {77–98},
publisher = {Brill},
title = {The {Arabic} dialect of a Šawāwī community of northern {Oman}},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Gardiner1904,
author = {Gardiner, Alan H.},
journal = {\textit{Zeitschrift} \textit{für} \textit{Ägyptische} \textit{Sprache} \
pages = {130–135},
title = {The word \textit{iwn3}},
volume = {41},
year = {1904}
}

\end{styleBibliography}
```

```
\begin{styleBibliography}
@book{Giannakidou1998,
address = {Amsterdam},
author = {Giannakidou, Anastasia},
publisher = {Benjamins},
title = {\textit{Polarity} \textit{sensitivity} \textit{as} \textit{(non)veridic
year = {1998}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Giannakidou2006,
address = {Oxford},
author = {Giannakidou, Anastasia},
booktitle = {\textit{The} \textit{{Blackwell}} \textit{Companion} \textit{to} \t
editor = {Martin Everaert and Henk Van Riemsdijk},
pages = {327–391},
publisher = {Blackwell},
title = {N-words and negative concord},
volume = {3},
year = {2006}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Hansen2013,
address = {Oxford},
author = {Hansen, Maj-Britt Mosegaard},
booktitle = {\textit{The} \textit{history} \textit{of} \textit{negation} \textit
editor = {David Willis, Christopher Lucas and Anne Breitbarth},
pages = {51–76},
publisher = {Oxford University Press},
title = {Negation in the history of {French}},
volume = {1},
year = {2013}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@article{Haspelmath2000,
author = {Haspelmath, Martin},
journal = {\textit{Linguistics}},
pages = {789–798},
title = {The relevance of extravagance: {{A}} reply to Bart Geurts},
volume = {38},
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{HaspelmathHaspelmath1996,
author = {Haspelmath, Martin and Josephine Caruana},
journal = {\textit{Rivista} \textit{di} \textit{Linguistica}},
pages = {213–227},
title = {Indefinite pronouns in Maltese},
volume = {8},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Heim1988,
address = {New York},
author = {Heim, Irene},
publisher = {Garland Publishing},
title = {\textit{The} \textit{semantics} \textit{of} \textit{definite} \textit{and}
year = {1988}
}

\end{styleBibliography}
```

```
\begin{styleBibliography}
@book{HeineHeine2005,
address = {Cambridge},
author = {Heine, Bernd and Tania Kuteva},
publisher = {Cambridge University Press},
title = {\textit{Language} \textit{contact} \textit{and} \textit{grammatical} \t
year = {2005}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Herin2016,
author = {Herin, Bruno},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}},
pages = {33–74},
title = {Elements of Domari dialectology},
volume = {23},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Herin2018,
address = {\textit{Arabic} \textit{in} \textit{contact}, 19–36. Amsterdam},
author = {Herin, Bruno},
publisher = {John Benjamins},
title = {The {Arabic} component in Domari},
year = {2018}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Holes2001,
address = {Leiden},
author = {Holes, Clive},
```

```
publisher = {Brill},
title = {\textit{Dialect,} \textit{culture,} \textit{and} \textit{society} \textit-
volume = {1},
year = {2001}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Holes2016,
address = {Leiden},
author = {Holes, Clive},
publisher = {Brill},
title = {\textit{Dialect,} \textit{culture,} \textit{and} \textit{society} \textit-
volume = {3},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Jespersen1917,
address = {Copenhagen},
author = {Jespersen, Otto},
note = {Høst.},
number = {1},
publisher = {A. F},
series = {Historisk-Filologiske Meddelelser},
title = {\textit{Negation} \textit{in} \textit{{English}} \textit{and} \textit{oth-
year = {1917}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Johnstone1987,
address = {London},
author = {Johnstone, T. M.},
```

```
publisher = {Routledge},
title = {\textit{Mehri} \textit{Lexicon} \textit{and} \textit{{English}—
Mehri} \textit{word-list}},
year = {1987}
}

\end{styleBibliography}

\begin{styleBibliography}
@thesis{Kahrel1996,
address = {Amsterdam},
author = {Kahrel, Peter},
school = {University of Amsterdam PhD},
title = {\textit{Aspects} \textit{of} \textit{negation}},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Kossmann2013,
address = {Leiden},
author = {Kossmann, Maarten},
publisher = {Brill},
title = {\textit{The} \textit{{Arabic}} \textit{influence} \textit{on} \textit{{
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Ladusaw1993,
address = {Berkeley},
author = {Ladusaw, William A.},
booktitle = {\textit{Proceedings} \textit{of} \textit{the} \textit{Nineteenth} \
editor = {Joshua S. Guenter, Barbara A. Kaiser and Cheryl C. Zoll},
pages = {437—46},
publisher = {Berkeley Linguistics Society},
```

```
title = {Negation, indefinites, and the {Jespersen} Cycle},
year = {1993}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui1996,
address = {Paris},
author = {Lafkioui, Mena},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{{
editor = {Salem Chaker and Dominique Caubet},
pages = {49--78},
publisher = {L'Harmattan},
title = {{La} négation en tarifit},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui2013,
address = {Berlin},
author = {Lafkioui, Mena},
booktitle = {\textit{{African}} \textit{{Arabic}:} \textit{Aprroaches} \textit{to}
editor = {Mena Lafkioui},
pages = {51--93},
publisher = {De Gruyter Mouton},
title = {Reinventing negation patterns in Moroccan {Arabic}},
year = {2013a}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui2013,
address = {Rome},
author = {Lafkioui, Mena},
```

```
booktitle = {\textit{Tilelli:} \textit{Scritti} \textit{in} \textit{onore} \text
editor = {Giorgio Francesco Arcodia, Federica Da Milano, Gabriele Iannàccaro and
pages = {113–130},
publisher = {Caissa Italia},
title = {Negation, grammaticalization and language change in {North} {Africa}: {
year = {2013b}
}

\end{styleBibliography}

\begin{styleBibliography}
@thesis{Laka1990,
address = {Cambridge, Mass.},
author = {Laka, Itziar},
school = {Massachusetts Institute of Technology PhD},
title = {\textit{Negation} \textit{in} \textit{syntax:} \textit{On} \textit{the}
year = {1990}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Lass1997,
address = {Cambridge},
author = {Lass, Roger},
publisher = {Cambridge University Press},
title = {\textit{Historical} \textit{linguistics} \textit{and} \textit{language}
year = {1997}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Lichtenberk1983,
address = {Honolulu},
author = {Lichtenberk, Frantisek},
publisher = {University of Hawaii Press},
title = {\textit{A} \textit{grammar} \textit{of} \textit{{Manam}}},
```

```
year = {1983}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Lucas2007,
author = {Lucas, Christopher},
journal = {\textit{Transactions} \textit{of} \textit{the} \textit{Philological} \te
note = {. doi:10.1111/j.1467-968X.2007.00189.x.},
pages = {398–431},
title = {{Jespersen}'s Cycle in {Arabic} and Berber},
volume = {105},
year = {2007}
}

\end{styleBibliography}

\begin{styleBibliography}
@misc{Lucas2009,
author = {Lucas, Christopher},
note = {University of Cambridge PhD.},
title = {\textit{The} \textit{development} \textit{of} \textit{negation} \textit{in
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lucas2013,
address = {Oxford},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{development} \textit{of} \textit{negation} \text
editor = {Anne Breitbarth, Christopher Lucas and David Willis},
pages = {399–452},
publisher = {Oxford University Press},
title = {Negation in the history of {Arabic} and Afro-Asiatic},
volume = {1},
```

```
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Lucas2018,
author = {Lucas, Christopher},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguist
pages = {45–71},
title = {On Wilmsen on the development of postverbal negation in dialectal {Arab
volume = {67},
year = {2018}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{LucasLucas2010,
author = {Lucas, Christopher and Elliott Lash},
journal = {\textit{Journal} \textit{of} \textit{Linguistics}}},
note = {. doi:10.1017/S9990235.},
pages = {379–413},
title = {Contact as catalyst: {{T}}he case for Coptic influence in the developme
urldate = {002222670},
volume = {46},
year = {2010}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Matras1999,
author = {Matras, Yaron},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}}},
pages = {1–58},
title = {The state of present-day Domari in Jerusalem},
volume = {11},
```

```
year = {1999}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Matras2007,
address = {Berlin},
author = {Matras, Yaron},
booktitle = {\textit{Grammatical} \textit{Borrowing} \textit{in} \textit{Cross-Ling
editor = {Yaron Matras and Jeanette Sakel},
pages = {151--164},
publisher = {Mouton de Gruyter},
title = {Grammatical borrowing in Domari},
year = {2007}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Matras2012,
address = {Berlin},
author = {Matras, Yaron},
publisher = {De Gruyter Mouton},
title = {\textit{A} \textit{grammar} \textit{of} \textit{Domari}},
year = {2012}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{McCarthyMcCarthy1964,
address = {Beirut},
author = {McCarthy, Richard Joseph and Faraj Raffouli},
publisher = {Librairie orientale},
title = {\textit{Spoken} \textit{{Arabic}} \textit{of} \textit{Baghdad:} \textit{Pa
year = {1964}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@article{Meillet1912,
author = {Meillet, Antoine},
journal = {\textit{Scientia}},
pages = {384--400},
title = {L'évolution des formes grammaticales},
volume = {12},
year = {1912}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Mughazy2003,
author = {Mughazy, Mustafa},
journal = {\textit{Journal} \textit{of} \textit{pragmatics}},
number = {8},
pages = {1143--1160},
title = {Metalinguistic negation and truth functions: {{T}}he case of {Egyptian}
volume = {35},
year = {2003}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Obler1990,
address = {Ann Arbor},
author = {Obler, Lorraine},
booktitle = {\textit{Studies} \textit{in} \textit{Near} \textit{{Eastern}} \text
editor = {James Bellamy},
pages = {132--152},
publisher = {University of Michigan},
title = {Reflexes of Classical {Arabic} \textit{šay'un} 'thing' in the modern di
year = {1990}
```

```
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Paradisi1961,
author = {Paradisi, Umberto},
journal = {\textit{Annali} \textit{dell'Istituto} \textit{Orientale} \textit{di} \t
pages = {79–91},
title = {Testi berberi di Augila (Cirenaïca)},
volume = {10},
year = {1961}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Pat-El2016,
author = {Pat-El, Na'ama},
journal = {\textit{Journal} \textit{of} \textit{Semitic} \textit{Studies}},
number = {1},
pages = {292–295},
sortname = {Pat-El, Na'ama},
title = {{David} Wilmsen, \textit{{Arabic}} \textit{Indefinites,} \textit{Interroga
volume = {61},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Rabhi1992,
author = {Rabhi, Allaoua},
journal = {\textit{Etudes} \textit{et} \textit{documents} \textit{berbères}},
pages = {139–145},
title = {Les particules de négation dans la Kabylie de l'Est},
volume = {9},
year = {1992}
```

```
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Rabhi1996,
address = {Paris},
author = {Rabhi, Allaoua},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \texti
editor = {Salem Chaker and Dominique Caubet},
pages = {23–34},
publisher = {L'Harmattan},
title = {De la négation en berbère: {{L}}es donnés algériennes},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Seeger2013,
address = {Wiesbaden},
author = {Seeger, Ulrich},
publisher = {Harrassowitz Verlag},
title = {\textit{Der} \textit{arabische} \textit{{Dialekt}} \textit{der} \texti
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Simeone-Senelle2003,
address = {1–13. Aden},
author = {Simeone-Senelle, Marie-Claude},
publisher = {University of Aden},
title = {Soqotri dialectology and the evaluation of the language endangerment},
year = {2003}
}
```

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Simeone-Senelle2011,
address = {Berlin},
author = {Simeone-Senelle, Marie-Claude},
booktitle = {\textit{The} \textit{{Semitic}} \textit{languages:} \textit{An} \text
editor = {Stefan Weninger, Geoffrey Khan, Michael P. Streck and Janet C. E. Watson}
pages = {1073–1113},
publisher = {Mouton de Gruyter},
title = {Modern {South} Arabian},
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Souag2009,
author = {Souag, Lameen},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguistik}
pages = {51–75},
title = {Siwa and its significance for {Arabic} dialectology},
volume = {51},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Souag2016,
author = {Souag, Lameen},
journal = {\textit{Linguistics}},
number = {1},
pages = {223–229},
title = {{Werner} Diem: \textit{Negation} \textit{in} \textit{{Arabic}:} \textit{A}
volume = {54},
year = {2016}
}

```
\end{styleBibliography}

\begin{styleBibliography}
@book{Souag2018,
address = {Amsterdam},
author = {Souag, Lameen},
publisher = {John Benjamins},
title = {{{Arabic}}-Berber-Songhay contact and the grammaticalization of "thing.
year = {2018}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{VanGelderen2018,
address = {The negative existential and other cycles},
author = {Van Gelderen, Elly},
note = {Unpublished manuscript. Arizona State University, ms. https://ling.auf.n
publisher = {Jespersen, Givón, and the copula cycle},
year = {2018}
}

\end{styleBibliography}

\begin{styleBibliography}
Wal Anonby, Christina van der. forthcoming. \textit{A} \textit{grammar} \textit{
\end{styleBibliography}

\begin{styleBibliography}
@article{Watson2011,
author = {Watson, Janet C. E.},
journal = {\textit{Salford} \textit{Working} \textit{Papers} \textit{in} \textit
pages = {27–40},
title = {{South} Arabian and Yemeni dialects},
volume = {1},
year = {2011}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@incollection{Willis2011,
address = {Berlin},
author = {Willis, David},
booktitle = {\textit{The} \textit{evolution} \textit{of} \textit{negation:} \texti
editor = {Pierre Larrivée and Richard Ingham},
pages = {285–323},
publisher = {De Gruyter},
title = {Negative polarity and the quantifier cycle: {{C}}omparative diachronic pe
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{WillisWillis2013,
address = {Vol. 1: Case studies. Oxford},
author = {Willis, David, Christopher Lucas and Anne Breitbarth},
publisher = {Oxford University Press},
title = {\textit{The} \textit{history} \textit{of} \textit{negation} \textit{in} \t
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Wilmsen2013,
author = {Wilmsen, David},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguistik}
pages = {5–31},
title = {The interrogative origin of the {{{Arabic}}} negator \textit{{}-š}: {{{E}}
volume = {58},
year = {2013}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@book{Wilmsen2014,
address = {Oxford},
author = {Wilmsen, David},
publisher = {Oxford University Press},
title = {\textit{{Arabic}} \textit{indefinites,} \textit{interrogatives} \textit
year = {2014}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Wilmsen2016,
author = {Wilmsen, David},
journal = {\textit{Folia} \textit{Orientalia}},
pages = {327–367},
title = {Another Croft cycle in {Arabic}: {{T}}he \textit{laysa} negative existe
volume = {53},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Auwera2009,
address = {Amsterdam},
author = {Auwera, Johan van der},
booktitle = {\textit{Cyclical} \textit{change}},
editor = {Elly Van Gelderen},
pages = {35–71},
publisher = {Benjamins},
title = {The {Jespersen} Cycles},
year = {2009}
}

\end{styleBibliography}
```

```
\begin{styleBibliography}
@article{AuweraAuwera2015,
author = {Auwera, Johan van der and Frens Vossen},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Southeast} \textit{As
pages = {24–38},
title = {Negatives between Chamic and Bahnaric},
volume = {8},
year = {2015}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{AuweraAuwera2016,
address = {Amsterdam},
author = {Auwera, Johan van der and Frens Vossen},
booktitle = {\textit{Cyclical} \textit{change} \textit{continued}},
editor = {Elly Van Gelderen},
pages = {189–218},
publisher = {Benjamins},
title = {{Jespersen} cycles in {Mayan}, Quechuan and Maipurean languages},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{AuweraAuwera2017,
author = {Auwera, Johan van der and Frens Vossen},
journal = {\textit{Linguistics} \textit{of} \textit{the} \textit{Tibeto-Burman} \te
number = {1},
pages = {40–58},
title = {{Kiranti} double negation: {{A}} copula conjecture},
volume = {40},
year = {2017}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@book{Behnstedt2016,
address = {(Trans.) Gwendolin Goldbloom. Leiden},
author = {Behnstedt, Peter},
publisher = {Brill},
title = {\textit{Dialect} \textit{atlas} \textit{of} \textit{{North}} \textit{{Y
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@misc{Bernabela2011,
author = {Bernabela, Roy S.},
note = {Leiden University M.A.},
title = {A phonology and morphology sketch of Šiħħi {Arabic} dialect of əlǦēdih
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{BerniniBernini1992,
address = {Bologna},
author = {Bernini, Giuliano and Paolo Ramat},
publisher = {Il Mulino},
title = {\textit{{La}} \textit{frase} \textit{negativa} \textit{nelle} \textit{l
year = {1992}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{BerniniBernini1996,
address = {Berlin},
author = {Bernini, Giuliano and Paolo Ramat},
```

```
publisher = {Mouton de Gruyter},
title = {\textit{Negative} \textit{sentences} \textit{in} \textit{the} \textit{lang
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Boumalk1996,
address = {Paris},
author = {Boumalk, Abdallah},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{
editor = {Salem Chaker and Dominique Caubet},
pages = {35–48},
publisher = {L'Harmattan},
title = {{La} négation en berbère marocain},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Brugnatelli1987,
address = {Milan},
author = {Brugnatelli, Vermondo},
booktitle = {\textit{{Atti}} \textit{della} \textit{4a} \textit{giornata} \textit{
editor = {Giuliano Bernini and Vermondo Brugnatelli},
pages = {53–62},
publisher = {Unicopli},
title = {{La} negazione discontinua in berbero e in arabo-magrebino},
year = {1987}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Buridant2000,
address = {Paris},
```

```
author = {Buridant, Claude},
publisher = {Sedes},
title = {\textit{Grammaire} \textit{nouvelle} \textit{de} \textit{l'ancien} \tex
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{CamilleriCamilleri2017,
address = {Stanford, CA},
author = {Camilleri, Maris and Louisa Sadler},
booktitle = {\textit{Proceedings} \textit{of} \textit{the} \textit{{LFG}'17} \te
editor = {Miriam Butt and Tracy Holloway King},
note = { http://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG
pages = {146–166},
publisher = {CSLI Publications},
title = {Negative sensitive indefinites in Maltese},
year = {2017}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Caubet1993,
address = {1: \textit{Phonologie} \textit{et} \textit{morphosyntaxe}. Paris},
author = {Caubet, Dominique},
publisher = {Peeters},
title = {\textit{L'arabe} \textit{marocain,} vol},
year = {1993a}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Caubet1993,
address = {2: \textit{Syntaxe} \textit{et} \textit{catégories} \textit{grammatic
author = {Caubet, Dominique},
```

```
publisher = {Peeters},
title = {\textit{L'arabe} \textit{marocain,} vol},
year = {1993b}
}

\end{styleBibliography}

\begin{styleBibliography}
Caubet, Dominique. 2011. Moroccan Arabic. In Lutz Edzard \& Rudolf de Jong (eds.),
\end{styleBibliography}

\begin{styleBibliography}
@incollection{Chaker1996,
address = {Paris},
author = {Chaker, Salem},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{e
editor = {Salem Chaker and Dominique Caubet},
pages = {9–22},
publisher = {L'Harmattan},
title = {Remarques préliminaires sur la négation en berbère},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{ChakerChaker1996,
address = {Paris},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{e
editor = {Chaker, Salem and Dominique Caubet},
publisher = {L'Harmattan},
title = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{et} \
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
```

```
@book{Coetsem1988,
address = {Dordrecht},
author = {Coetsem, Frans van},
publisher = {Foris},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two}
year = {1988}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Coetsem2000,
address = {Heidelberg},
author = {Coetsem, Frans van},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theor
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Cowell1964,
address = {Washington DC},
author = {Cowell, Mark W.},
publisher = {Georgetown University Press},
title = {\textit{A} \textit{reference} \textit{grammar} \textit{of} \textit{Syri
year = {1964}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Croft1991,
author = {Croft, William},
journal = {\textit{Journal} \textit{of} \textit{Linguistics}},
number = {1},
pages = {1–27},
```

```
title = {The evolution of negation},
volume = {27},
year = {1991}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Dahl1979,
author = {Dahl, Östen},
journal = {\textit{Linguistics}},
pages = {79--106},
sortname = {Dahl, Osten},
title = {Typology of sentence negation},
volume = {17},
year = {1979}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{DevosDevos2013,
author = {Devos, Maud and Johan van der Auwera},
journal = {\textit{Journal} \textit{of} \textit{African} \textit{Languages} \texti
pages = {205--274},
title = {{Jespersen} cycles in {Bantu}: {{D}}ouble and triple negation},
volume = {34},
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Diem2014,
address = {Wiesbaden},
author = {Diem, Werner},
publisher = {Harrassowitz},
title = {\textit{Negation} \textit{in} \textit{{Arabic}:} \textit{A} \textit{study}
```

```
year = {2014}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Eades2009,
address = {Leiden},
author = {Eades, Domenyk},
booktitle = {\textit{{Arabic}} \textit{dialectology} \textit{in} \textit{honour}
editor = {Enam Al-Wer and Rudolf De Jong},
pages = {77–98},
publisher = {Brill},
title = {The {Arabic} dialect of a Šawāwī community of northern {Oman}},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Gardiner1904,
author = {Gardiner, Alan H.},
journal = {\textit{Zeitschrift} \textit{für} \textit{Ägyptische} \textit{Sprache}
pages = {130–135},
title = {The word \textit{iwn3}},
volume = {41},
year = {1904}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Giannakidou1998,
address = {Amsterdam},
author = {Giannakidou, Anastasia},
publisher = {Benjamins},
title = {\textit{Polarity} \textit{sensitivity} \textit{as} \textit{(non)veridic
year = {1998}
```

```
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Giannakidou2006,
address = {Oxford},
author = {Giannakidou, Anastasia},
booktitle = {\textit{The} \textit{{Blackwell}} \textit{Companion} \textit{to} \text
editor = {Martin Everaert and Henk Van Riemsdijk},
pages = {327–391},
publisher = {Blackwell},
title = {N-words and negative concord},
volume = {3},
year = {2006}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Hansen2013,
address = {Oxford},
author = {Hansen, Maj-Britt Mosegaard},
booktitle = {\textit{The} \textit{history} \textit{of} \textit{negation} \textit{i
editor = {David Willis, Christopher Lucas and Anne Breitbarth},
pages = {51–76},
publisher = {Oxford University Press},
title = {Negation in the history of {French}},
volume = {1},
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Haspelmath2000,
author = {Haspelmath, Martin},
journal = {\textit{Linguistics}},
```

```
pages = {789–798},
title = {The relevance of extravagance: {{A}} reply to Bart Geurts},
volume = {38},
year = {2000}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{HaspelmathHaspelmath1996,
author = {Haspelmath, Martin and Josephine Caruana},
journal = {\textit{Rivista} \textit{di} \textit{Linguistica}},
pages = {213–227},
title = {Indefinite pronouns in Maltese},
volume = {8},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Heim1988,
address = {New York},
author = {Heim, Irene},
publisher = {Garland Publishing},
title = {\textit{The} \textit{semantics} \textit{of} \textit{definite} \textit{a
year = {1988}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{HeineHeine2005,
address = {Cambridge},
author = {Heine, Bernd and Tania Kuteva},
publisher = {Cambridge University Press},
title = {\textit{Language} \textit{contact} \textit{and} \textit{grammatical} \t
year = {2005}
```

```
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Holes2001,
address = {Leiden},
author = {Holes, Clive},
publisher = {Brill},
title = {\textit{Dialect,} \textit{culture,} \textit{and} \textit{society} \textit
volume = {1},
year = {2001}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Holes2016,
address = {Leiden},
author = {Holes, Clive},
publisher = {Brill},
title = {\textit{Dialect,} \textit{culture,} \textit{and} \textit{society} \textit
volume = {3},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Jespersen1917,
address = {(Historisk-Filologiske Meddelelser 1). Copenhagen},
author = {Jespersen, Otto},
note = {Høst.},
publisher = {A. F},
title = {\textit{Negation} \textit{in} \textit{{English}} \textit{and} \textit{oth
year = {1917}
}
```

```
\end{styleBibliography}

\begin{styleBibliography}
@book{Johnstone1987,
address = {London},
author = {Johnstone, T. M.},
publisher = {Routledge},
title = {\textit{Mehri} \textit{Lexicon} \textit{and} \textit{{English}—
Mehri} \textit{word-list}},
year = {1987}
}

\end{styleBibliography}

\begin{styleBibliography}
@thesis{Kahrel1996,
address = {Amsterdam},
author = {Kahrel, Peter},
school = {University of Amsterdam PhD},
title = {Aspects of negation},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Kossmann2013,
address = {Leiden},
author = {Kossmann, Maarten},
publisher = {Brill},
title = {\textit{The} \textit{{Arabic}} \textit{influence} \textit{on} \textit{{
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Ladusaw1993,
```

```
address = {Berkeley},
author = {Ladusaw, William A.},
booktitle = {\textit{Proceedings} \textit{of} \textit{the} \textit{Nineteenth} \tex
editor = {Joshua S. Guenter, Barbara A. Kaiser and Cheryl C. Zoll},
pages = {437–46},
publisher = {Berkeley Linguistics Society},
title = {Negation, indefinites, and the {Jespersen} Cycle},
year = {1993}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui1996,
address = {Paris},
author = {Lafkioui, Mena},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \textit{e
editor = {Salem Chaker and Dominique Caubet},
pages = {49–78},
publisher = {L'Harmattan},
title = {{La} négation en tarifit},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui2013,
address = {Berlin},
author = {Lafkioui, Mena},
booktitle = {\textit{{African}} \textit{{Arabic}:} \textit{Aprroaches} \textit{to}
editor = {Mena Lafkioui},
pages = {51–93},
publisher = {De Gruyter Mouton},
title = {Reinventing negation patterns in Moroccan {Arabic}},
year = {2013a}
}
```

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lafkioui2013,
address = {Rome},
author = {Lafkioui, Mena},
booktitle = {\textit{Tilelli:} \textit{Scritti} \textit{in} \textit{onore} \text
editor = {Giorgio Francesco Arcodia, Federica Da Milano, Gabriele Iannàccaro and
pages = {113–130},
publisher = {Caissa Italia},
title = {Negation, grammaticalization and language change in {North} {Africa}: {
year = {2013b}
}

\end{styleBibliography}

\begin{styleBibliography}
@thesis{Laka1990,
address = {Cambridge, Mass.},
author = {Laka, Itziar},
school = {Massachusetts Institute of Technology PhD},
title = {Negation in syntax: {{O}}n the nature of functional categories and proj
year = {1990}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Lass1997,
address = {Cambridge},
author = {Lass, Roger},
publisher = {Cambridge University Press},
title = {\textit{Historical} \textit{linguistics} \textit{and} \textit{language}
year = {1997}
}

\end{styleBibliography}

```
\begin{styleBibliography}
@book{Lichtenberk1983,
address = {Honolulu},
author = {Lichtenberk, Frantisek},
publisher = {University of Hawaii Press},
title = {\textit{A} \textit{grammar} \textit{of} \textit{{Manam}}},
year = {1983}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Lucas2007,
author = {Lucas, Christopher},
journal = {\textit{Transactions} \textit{of} \textit{the} \textit{Philological} \te
note = {. doi:10.1111/j.1467-968X.2007.00189.x.},
pages = {398–431},
title = {{Jespersen}'s Cycle in {Arabic} and Berber},
volume = {105},
year = {2007}
}

\end{styleBibliography}

\begin{styleBibliography}
@misc{Lucas2009,
author = {Lucas, Christopher},
note = {University of Cambridge PhD.},
title = {The development of negation in {Arabic} and Afro-Asiatic},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Lucas2013,
address = {Oxford},
author = {Lucas, Christopher},
```

```
booktitle = {\textit{The} \textit{development} \textit{of} \textit{negation} \te
editor = {Anne Breitbarth, Christopher Lucas and David Willis},
pages = {399–452},
publisher = {Oxford University Press},
title = {Negation in the history of {Arabic} and Afro-Asiatic},
volume = {1},
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Lucas2018,
author = {Lucas, Christopher},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguist
pages = {45–71},
title = {On Wilmsen on the development of postverbal negation in dialectal {Arab
volume = {67},
year = {2018}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{LucasLucas2010,
author = {Lucas, Christopher and Elliott Lash},
journal = {\textit{Journal} \textit{of} \textit{Linguistics}},
note = {. doi:10.1017/S9990235.},
pages = {379–413},
title = {Contact as catalyst: {{T}}he case for Coptic influence in the developme
urldate = {002222670},
volume = {46},
year = {2010}
}

\end{styleBibliography}

\begin{styleBibliography}
```

@article{Matras1999,
author = {Matras, Yaron},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}},
pages = {1–58},
title = {The state of present-day Domari in Jerusalem},
volume = {11},
year = {1999}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Matras2007,
address = {Berlin},
author = {Matras, Yaron},
booktitle = {\textit{Grammatical} \textit{Borrowing} \textit{in} \textit{Cross-Lin
editor = {Yaron Matras and Jeanette Sakel},
pages = {151–164},
publisher = {Mouton de Gruyter},
title = {Grammatical borrowing in Domari},
year = {2007}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Matras2012,
address = {Berlin},
author = {Matras, Yaron},
publisher = {De Gruyter Mouton},
title = {\textit{A} \textit{grammar} \textit{of} \textit{Domari}},
year = {2012}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{McCarthyMcCarthy1964,

```
address = {Beirut},
author = {McCarthy, Richard Joseph and Faraj Raffouli},
publisher = {Librairie orientale},
title = {\textit{Spoken} \textit{{Arabic}} \textit{of} \textit{Baghdad:} \textit
year = {1964}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Meillet1912,
author = {Meillet, Antoine},
journal = {\textit{Scientia}},
pages = {384–400},
title = {L'évolution des formes grammaticales},
volume = {12},
year = {1912}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Mughazy2003,
author = {Mughazy, Mustafa},
journal = {\textit{Journal} \textit{of} \textit{pragmatics}},
number = {8},
pages = {1143–1160},
title = {Metalinguistic negation and truth functions: {{T}}he case of {Egyptian}
volume = {35},
year = {2003}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Obler1990,
address = {Ann Arbor},
author = {Obler, Lorraine},
```

```
booktitle = {\textit{Studies} \textit{in} \textit{Near} \textit{{Eastern}} \textit-
editor = {James Bellamy},
pages = {132–152},
publisher = {University of Michigan},
title = {Reflexes of Classical {Arabic} \textit{šay'un} 'thing' in the modern dial
year = {1990}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Paradisi1961,
author = {Paradisi, Umberto},
journal = {\textit{Annali} \textit{dell'Istituto} \textit{Orientale} \textit{di} \
pages = {79–91},
title = {Testi berberi di Augila (Cirenaïca)},
volume = {10},
year = {1961}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Pat-El2016,
author = {Pat-El, Na'ama},
journal = {\textit{Journal} \textit{of} \textit{Semitic} \textit{Studies}},
number = {1},
pages = {292–295},
sortname = {Pat-El, Na'ama},
title = {{David} Wilmsen, \textit{{Arabic}} \textit{Indefinites,} \textit{Interroga
volume = {61},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Rabhi1992,
```

```
author = {Rabhi, Allaoua},
journal = {\textit{Etudes} \textit{et} \textit{documents} \textit{berbères}},
pages = {139–145},
title = {Les particules de négation dans la Kabylie de l'Est},
volume = {9},
year = {1992}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Rabhi1996,
address = {Paris},
author = {Rabhi, Allaoua},
booktitle = {\textit{{La}} \textit{négation} \textit{en} \textit{berbère} \texti
editor = {Salem Chaker and Dominique Caubet},
pages = {23–34},
publisher = {L'Harmattan},
title = {De la négation en berbère: {{L}}es donnés algériennes},
year = {1996}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Seeger2013,
address = {Wiesbaden},
author = {Seeger, Ulrich},
publisher = {Harrasowitz Verlag},
title = {\textit{Der} \textit{arabische} \textit{{Dialekt}} \textit{der} \textit
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Simeone-Senelle2003,
address = {1–13. Aden},
```

```
author = {Simeone-Senelle, Marie-Claude},
publisher = {University of Aden},
title = {Soqotri dialectology and the evaluation of the language endangerment},
year = {2003}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Simeone-Senelle2011,
address = {Berlin},
author = {Simeone-Senelle, Marie-Claude},
booktitle = {\textit{The} \textit{{Semitic}} \textit{languages:} \textit{An} \text
editor = {Stefan Weninger, Geoffrey Khan, Michael P. Streck and Janet C. E. Watson
pages = {1073–1113},
publisher = {Mouton de Gruyter},
title = {Modern {South} Arabian},
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Souag2009,
author = {Souag, Lameen},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguistik
pages = {51–75},
title = {Siwa and its significance for {Arabic} dialectology},
volume = {51},
year = {2009}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Souag2016,
author = {Souag, Lameen},
journal = {\textit{Linguistics}},
```

number = {1},
pages = {223–229},
title = {{Werner} Diem: \textit{Negation} \textit{in} \textit{{Arabic}:} \textit
volume = {54},
year = {2016}
}

\end{styleBibliography}

\begin{styleBibliography}
Wal Anonby, Christina van der. forthcoming. A grammar of Kumzari: A mixed Perso-
\end{styleBibliography}

\begin{styleBibliography}
@article{Watson2011,
author = {Watson, Janet CE},
journal = {\textit{Salford} \textit{Working} \textit{Papers} \textit{in} \textit
pages = {27–40},
title = {{South} Arabian and Yemeni dialects},
volume = {1},
year = {2011}
}

\end{styleBibliography}

\begin{styleBibliography}
@incollection{Willis2011,
address = {Berlin},
author = {Willis, David},
booktitle = {\textit{The} \textit{evolution} \textit{of} \textit{negation:} \tex
editor = {Pierre Larrivée and Richard Ingham},
pages = {285–323},
publisher = {De Gruyter},
title = {Negative polarity and the quantifier cycle: {{C}}omparative diachronic
year = {2011}
}

\end{styleBibliography}

```
\begin{styleBibliography}
@book{WillisWillis2013,
address = {. Vol. 1: Case studies. Oxford},
author = {Willis, David, Christopher Lucas and Anne Breitbarth},
publisher = {Oxford University Press},
title = {\textit{The} \textit{history} \textit{of} \textit{negation} \textit{in} \t
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Wilmsen2013,
author = {Wilmsen, David},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguistik}
pages = {5–31},
title = {The interrogative origin of the {{{Arabic}}} negator \textit{{}-š}: {{{E}}
volume = {58},
year = {2013}
}

\end{styleBibliography}

\begin{styleBibliography}
@book{Wilmsen2014,
address = {Oxford},
author = {Wilmsen, David},
publisher = {Oxford University Press},
title = {\textit{{Arabic}} \textit{indefinites,} \textit{interrogatives} \textit{a
year = {2014}
}

\end{styleBibliography}

\begin{styleBibliography}
@article{Wilmsen2016,
author = {Wilmsen, David},
```

```
journal = {\textit{Folia} \textit{Orientalia}},
pages = {327–367},
title = {Another Croft cycle in {Arabic}: {{T}}he \textit{laysa} negative existe
volume = {53},
year = {2016}
}
```

# Chapter 8

# Maltese

## Christopher Lucas
SOAS, University of London

## Slavomír Čéplö
Institute of Oriental Studies, Slovak Academy of Sciences / Abteilung für Byzanzforschung, Österreichische Akademie der Wissenschaften

This chapter presents an overview of the most prominent contact-induced developments in the history of Maltese, a language which is genetically a variety of Arabic, but which has undergone significant changes, largely as a result of lengthy contact with Sicilian, Italian, and English. We first address the precise affiliation of Maltese and the nature of the historical and ongoing contact situations, before detailing relevant developments in the realms of phonology, inflectional and derivational morphology, syntax and lexicon.

## 1 Maltese and Arabic

From a historical point of view, Maltese is a variety of spoken Arabic, albeit one that has undergone far-reaching changes as a result of sustained and intensive contact with Italo-Romance varieties, and more recently also with English. This is a fact about which there is no controversy among contemporary linguists. It should be noted, however, that a mix of social, cultural, historical, political, and indeed linguistic factors has led to a situation in which many Maltese people today view their language as Semitic, but not a type of Arabic. Since we are concerned here only with the historical perspective, we will not dwell on the vexed question of whether or not contemporary Maltese should be classified as an "Arabic dialect".[1] Suffice it to say that the idea, first popularized by **desoldanis1750**

---

[1]Note that Maltese itself has a number of different dialects, one of which – that of the major towns, and the variety used in media, literature and administration – is referred to as standard

and **vassalli1791**, that Maltese is a variety of Phoenician or Punic, has been shown since at least since **gesenius1810** and **desacy1829** to be entirely without merit.

Since the Phoenicians and then the Carthaginians occupied Malta for much of the first millennium BCE, followed by Roman and Byzantine occupation for much of the first millennium CE, it would seem *prima facie* likely that elements of the languages of these occupiers would survive into contemporary Maltese. **brincat1995** shows, however, based on the account of al-Ḥimyarī, that Malta was to all intents and purposes uninhabited in the period between its conquest by the Arabs in 870 CE and the first concerted efforts at colonization by Arabic-speaking Muslims in 1048–9 CE. It is for this reason that the Semitic component of Maltese phonology, morphology, syntax and lexicon is Arabic and Arabic only (see also **grech1961**).

As for the provenance of the Arabic component of contemporary Maltese, there is no doubt that the most important source is a variety of Maghrebi (Western) Arabic. This is evident from grammatical features such as: the pan-Maghrebi extension to the singular of the first-person *n-* prefix of the imperfect verb paradigm (see Table 1); the loss of a gender distinction in the second person singular, in pronouns and both perfect and imperfect verbs, as in urban Tunisian Arabic varieties (**gibson2011**); variable rearticulation of the definite article on postnominal adjectives in definite noun phrases, as in (1), found also in Casablanca Arabic (**harrell2004**); and the *-il* suffix of the numerals 'eleven' to 'nineteen' in determiner use, as in (2), which also occurs in the Arabic dialects of Casablanca (**caubet2011**) and Tlemcen (**tainecheikh2011**).[2]

Table 1: First person imperfect 'write' in Eastern and Western Arabic

|  | Eastern | | Western | |
|---|---|---|---|---|
|  | Classical Arabic | Baghdad Arabic | Casablanca Arabic | Maltese |
| Singular | *aktub* | *aktib* | *nəktəb* | *nikteb* |
| Plural | *naktub* | *niktib* | *nkətbu* | *niktbu* |

---

Maltese. Except where specified, this chapter deals exclusively with the standard variety of Maltese.

[2]Unless otherwise specified, all numbered examples present data from Maltese. All Maltese examples in this chapter are rendered using standard Maltese orthography.

(1)  il-kelb    (l-)abjad
     DEF-dog (DEF)-white
     'the white dog' (**gattinpress**)

(2)  it-tnax-**il**    appostlu
     DEF-twelve-IL apostle
     'the twelve apostles'

Narrowing matters down further, Zammit's (**zammit2014**) study of lexicon shared between Maltese and the Arabic dialect of Sfax offers yet more support (see also **vanhove1998**) for the geographically unsurprising conclusion that Maltese is more closely related to the traditional (so-called pre-Hilali; see Benkato, this volume) urban Tunisian dialects than to any other extant Arabic variety. This is not to suggest, however, that the Arabic component of Maltese resembles these dialects in all respects. **borg1996** lists a number of areas in which Maltese accords more closely with Levantine Arabic dialects than with those of the Maghreb. But the social and political history of Malta after the end of direct Arab rule in 1127 CE is such that most or all of these similarities should be understood as the failure of Maltese to participate in innovations that later spread through the mainland Maghrebi varieties, and not as evidence of influence of Eastern Arabic on the formation of Maltese.

## 2 Contact with Italo-Romance and English

### 2.1 Italo-Romance

A comprehensive history of immigration to Malta in the medieval period is yet to be written (if indeed such a history is possible at all, given the apparently scarce documentary evidence). It is therefore impossible to give precise details of the sociolinguistic conditions under which the Arabic variety spoken in Malta came into contact with varieties of Italo-Romance in the course of the second millennium. We can, however, sketch the broad outlines of this process, and make some reasonable inferences.

The Arabic-speaking settlers who colonized Malta in 1048–9 CE can be assumed to have come from either Sicily or southern Italy or both (**brincat1995**), but in any case it seems likely that at least some of these came speaking a variety of Sicilian in addition to Arabic. Even after Malta was brought under Norman control in 1127 CE by Roger II of Sicily, and went on to be part of the Kingdom of Sicily, there does not, however, seem to have been really large-scale immigration

of non-Arabic speakers to Malta at any point, a fact which is of course consistent with the survival of the Maltese language until today. Unsurprisingly from a geographical and political perspective, what immigration there was appears to have come overwhelmingly from Sicily and southern Italy, with lesser numbers coming also from Spain (**ballou1893**; **blouet1967**; **fiorini1986**; **goodwin2002**).

Comprising mostly soldiers, craftsmen and churchmen of various types, it would appear that this immigration was disproportionately male. In addition to families in which the only language spoken was Maltese, there must, therefore, have been significant numbers of families in medieval Malta in which the father spoke only Sicilian natively and the mother spoke only Maltese natively, with communication necessarily involving second-language speech by one or both parents. Children of such families would therefore have been exposed minimally to native and non-native Maltese speech and native Sicilian speech.

From the perspective of Van Coetsem's (**VanCoetsem1988**; **coetsem2000**) framework for understanding contact-induced change, therefore, it seems highly likely that transfer from Sicilian to Maltese occurred both through imposition under source-language agentivity (by L1 Sicilian speakers) and borrowing under recipient-language agentivity (by L1 Maltese speakers).

There is no doubt that, alongside Sicilian, (Tuscan) Italian had an important place in Maltese life over many centuries, starting at the latest in 1530, when it became the official language of government under the regime of the Knights of Malta. But as Comrie & Spagnol (**comriespagnol2016**) point out, Italian did not gain a foothold at the expense of Sicilian among bilingual Maltese until the later eighteenth century, and given its social function as a vehicle for government, education and high culture, rather than the native language of a significant proportion of ordinary Maltese, it is reasonable to say that transfer from Italian will have been mediated predominantly by borrowing under recipient-language agentivity.

## 2.2 English

Starting in 1800, when Malta became a protectorate of the British Empire, English gradually began to supplant Italian as the language of government, education and high culture, being joined in that role by the Maltese language itself only in the last few decades. English is now widely spoken in Malta: according to 2011 census data (**census2011**), 94.6% of the population of Malta reported speaking Maltese "well" or "average[ly]", while 82.1% reported the same for English. English is a native language for only a very small percentage of Maltese residents, however: **scirihavassallo2006** put the figure at 2%. As with Italian, then, transfer from

English to Maltese will overwhelmingly have occurred through borrowing under recipient-language agentivity. With the Maltese variety of English, the reverse is true of course: here the transfer from English to Maltese will have been almost exclusively imposition under source-language agentivity by native speakers of Maltese, resulting in such hallmark features of Maltese English as word-final obstruent devoicing (cf. §3.1.2 below), and the use of *but* in clause-final position (**lucas2015**).

Given that transfer from English was and is restricted to borrowing in Van Coetsem's sense, while the more extensive and long-lasting contact with Sicilian will have involved both borrowing and imposition, it is not surprising that a picture will emerge in the following sections whereby Italo-Romance dominates as a source of contact-induced changes across all linguistic domains, with English playing a much more modest role, largely restricted to lexicon and associated inflectional morphology.

# 3 Phonology[3]

## 3.1 Consonants

### 3.1.1 Additions to the native phonemic inventory

One of the most salient, and uncontroversially contact-induced, innovations within Maltese phonology is the addition of at least five (arguably seven) consonant phonemes. This came about through the transfer (presumably borrowing) of Italo-Romance and English lexical items without subsequent adaptation to the original native inventory (compare, e.g., Maltese *pulizija* with unadapted initial [p] and Cairene Arabic *bulīs* 'police'). The five uncontroversial additions are /p/, /v/, /t͡s/, /t͡ʃ/ and /g/ (orthographically: <p>, <v>, <z>, <ċ> and <g>; see Table 2), as in *evaporazzjoni* 'evaporation' and *granċ* 'crab'. One can also make a case for an innovative borrowed phoneme /d͡z/. There are no minimal pairs demonstrating a phonemic distinction between /d͡z/ and /t͡s/ (and both are represented by <z> in the orthography), but **borgazzopardimaltese** point out that /d͡z/ occurs in environments not requiring a voiced obstruent, as in *gazzetta* /gɐˈd͡zːɛtːɐ/ 'newspaper'. More marginal is /ʒ/, which **mifsud2011** and **borgazzopardimaltese** point out can be found in recent loanwords from English, such as *televixin* 'television' and *bex* 'beige', though whether all speakers voice the <x> in these items is uncertain.

---

[3]**For useful overviews of the phonology of Maltese, see borg1997 and cohen1966.**

Table 2: Inventory of consonants. Symbols are Maltese orthography.

| | Labial | Alveolar | Post-alveolar | Palatal | Velar | Laryngeal |
|---|---|---|---|---|---|---|
| Plosive | p | t | | | k | q |
| | b | d | | | g | |
| Affricate | | z | ċ | | | |
| | | | ġ | | | |
| Fricative | f | s | x | | | ħ |
| | v | ż | | | | |
| Nasal | m | n | | | | |
| Trill | | r | | | | |
| Lateral | | l | | | | |
| Approximant | w | | | j | | |

Proto-Semitic *g, represented as <ج> in Arabic script, and usually rendered /dʒ/ when standard Arabic is spoken, is reflected as /dʒ/ (orthographic <ġ>) in Maltese. This appears to be a retention of the original Maghrebi realization of this phoneme, other Maghrebi varieties having in general deaffricated it to /ʒ/ (cf. **heath2002**). Unlike some other Maghrebi varieties, however, the Maltese reflex of <ج> does not become /g/ before sibilants (cp. Maltese *ġewż* vs. Casablanca *gūz* 'walnuts').[4] Similarly, Proto-Semitic *q (on which more below), is never reflected as /g/ (orthographic <g>) in Maltese (cf. **vanhove1998**), meaning that the presence of /g/ in the Maltese phonemic inventory is certainly due to its occurence in numerous lexical borrowings. The majority of these are from Italo-Romance (e.g. *gwerra* 'war'), but some are from Berber (e.g. *gendus* 'calf' < Berber *agenduz*, **naitzerrad2002**), suggesting that /g/ as an independent phoneme has been present in Maltese since the earliest days of Arabic speech on the islands.[5]

### 3.1.2 Losses, mergers and shifts

Alongside these additions, the Maltese consonant phoneme inventory has also witnessed a number of losses and mergers. Clearly it is not possible to establish with certainty whether or not these changes were due to contact, but various

---

[4]An exception is *gżira* 'island' < Arabic *jazīra*, perhaps to be explained by direct contiguity with the sibilant.

[5]There are also some sporadic examples of /g/ < *k in Arabic roots, e.g. *gideb* 'to lie'. See **cohen1966** for further details.

considerations make it reasonable to assume that contact at least accelerated these changes. For example, the inherited emphatic (pharyngealized/uvularized) consonants – *ṣ, *ṭ, *ḍ, and *ð̣ – have all merged with their non-emphatic counterparts, as in *sħab* /sħɐːb/ 'clouds' < *saḥāb*, and also 'companions' < *ʔaṣḥāb*. Note in this connection that among other Arabic varieties, it is only a handful of those most strongly affected by contact (such as pidgins and creoles, as well as Cypriot Maronite Arabic; see Avram, this volume; Walter, this volume) that have merged the emphatic consonants in this way. This suggests that non-native acquisition of Maltese by Italo-Romance speakers precipitated this change (i.e. that it involves SL-agentivity in Van Coetsem's **VanCoetsem1988**; **coetsem2000** terms).

In addition to the loss of the emphatic consonants, Maltese has undergone significant losses and mergers among the velar and laryngeal phonemes.

Perhaps most saliently, an earlier version of what is today standard Maltese merged and then lost the voiced uvular/velar fricative *ɣ and the voiced pharyngeal fricative *ʕ. In Maltese's rather etymologizing orthography, these historic phonemes are given the digraph symbol <għ>. In general, this symbol either has no phonetic correlate, as in *għajn* /ɐjn/ 'eye, spring' and *għasfur* /ɐsˈfuːr/ 'bird', or otherwise corresponds to the lengthening of a vowel in morphological patterns where the vowel would ordinarily be short, as in the stem I *CaCeC* verb *għamel* /ˈɐːmɛl/ 'to do'. That the two original phonemes first merged and were then lost in standard Maltese can be inferred from the behaviour of <għ> + <h> sequences. These are realised as /ħː/ in roots where <għ> reflects *ʕ (e.g. *semagħ-ha* /sɛˈmɐħːɐ/ hear.PRF.3SG.M-3SG.F, 'he heard it'), where other Arabic varieties behave similarly (cf. **woidich2006**), but also, unlike other Arabic varieties, in roots where <għ> reflects *ɣ (e.g. *ferragħ-ha* /fɛrˈrɐħːɐ/ pour.PRF.3SG.M-3SG.F, 'he poured it out' < Arabic *farraɣ* 'to empty'). This merger and subsequent loss did not take place in all varieties of Maltese. To this day, there are apparently speakers of dialectal Maltese whose speech preserves both *ɣ as a velar fricative, and *ʕ as a pharyngeal fricative (**klimiuk2017**). The fact that the merger and loss of these two phonemes is more advanced in the standard language of the major conurbations and less so in the dialects of more isolated villages suggests that contact-induced change played an important role here, with non-native speakers of Maltese presumably being the principal agents of change.

Arguably the most interesting set of mergers and losses concerns the voiceless fricatives, which represent a case of considerable phonemic reorganization despite relatively little change at the phonetic level. The phonemic changes in this domain are as follows. First, *h, while maintained in the orthography (as <h>), has merged with /ħ/ in codas (e.g. *ikrah* /ɪkˈrɐħ/ 'ugly') and sporadically in

onsets (e.g. *naħaq* /nɐˈħɐʔ/ < *nahaq* 'to bray (of donkeys)'), and is otherwise lost altogether (e.g. *hemm* /ɛm:/ 'there'). The Maltese phoneme /ħ/ thus represents the continuation of the voiceless pharyngeal fricative *ḥ, as well as the partial merger of *h. Moreover, original *ḫ, the voiceless uvular/velar fricative, has also merged with /ħ/, as in *ħajt* 'thread' < *ḫayṭ*, and also 'wall' < *ḫāyiṭ*. Strikingly, however, the single Maltese phoneme /ħ/ exhibits considerable inter- and intra-speaker variation in its precise realization, such that glottal, pharyngeal, and velar/uvular voiceless fricative realizations may commonly be heard (**borgazzopardimaltese**), and it is in this sense there has been little phonetic change despite the considerable phonological reorganization.

Like the loss of the emphatic consonants, the loss or merger of *h, as well as one or more of the pharyngeal and velar/uvular fricatives, is restricted to a handful of Arabic varieties that have been very strongly affected by contact (see, e.g., Walter, this volume). As such, these changes too are suggestive of imposition by non-native speakers lacking these sounds in their native phonemic inventory (as was the case for speakers of the Romance varieties with which Maltese has had the most intense contact, cf. **loporcaro2011**). On the other hand, the preservation of the glottal and pharyngeal fricatives as allophones of /ħ/ complicates this picture, such that the role of contact in bringing about these particular changes must remain uncertain for now.

It is similarly hard to diagnose the causes of the shift of *q to glottal stop (nevertheless written as <q> in Maltese orthography) and the stopping of the interdental fricatives *θ and *ð. In both cases, however, we can at least rule out with confidence any suggestion that these are ancient changes that predate the arrival of Arabic in Malta, or are historically connected to similar realizations in the Arabic dialects of urban centres in the Maghreb, Egypt, and the Levant. Written records of earlier Maltese clearly show that a dorsal realization of *q, and an interdental fricative realization of *θ and *ð, survived until at least the late eighteenth century (**avram2012**; **avram2014**). It is at least plausible, therefore, that contact with Italo-Romance played a role in these changes too, but firm evidence on this point is so far lacking.

Finally, a well-known feature of contemporary Maltese (and Maltese English) phonology is the devoicing of word-final obstruents, as in *ħadd* [ħɐt:] 'nobody'. **avram2017** shows that devoicing gradually diffused across the Maltese lexicon over the course of about two centuries from the late 16th century onwards, and he makes a strong case that the initial trigger for this development was imposition by native speakers of Sicilian and Italian, since word-final obstruent devoicing has been shown by various studies (e.g. **flegemunromackay1995**) to be a frequent

feature of the L2 speech of L1 speakers of Romance languages.

## 3.2 Vowels

Maltese has a much richer vowel phoneme inventory than typical Maghrebi Arabic dialects, with, among the monophthongs, five short-vowel qualities /ɪ, ɛ, ɐ, ɔ, ʊ/ (orthographic <i, e, a, o, u>), and six long-vowel qualities /iː, ɪː, ɛː, ɐː, ɔː, uː/ (orthographic <i, ie, e, a, o, u>), as well as seven distinct diphthongs (with a number of different orthographies – see **borgazzopardimaltese** for details): /ɪʊ, ɛɪ, ɛʊ, ɐɪ, ɐʊ, ɔɪ, ɔʊ/. Compare this with the three-vowel-quality system of Tunis Arabic, which also lacks diphthongs (**gibson2011**).

Since the Italo-Romance languages have vowel systems of a similar richness to Maltese, one might assume that this proliferation of vowel phonemes is a straightforward case of transfer. This is, in general, not the case, however. The majority of new phonemic distinctions are at least partially the result of the loss of emphatic consonants and of *ʕ,[6] which led to the phonemicization of vowel qualities that were previously merely allophonic. Note also that the innovative lax close front long vowel /ɪː/ is apparently an entirely internal development – the outcome of an extreme raising of the front allophone of *ā (so-called *imāla*), as in *ktieb* /ktɪːb/ 'book' < *kitāb*.

Following Krier (**krier1976**), we can nevertheless point to three innovations in this domain which do seem to be the direct result of lexical borrowing from Italo-Romance.

Krier (**krier1976**) points out first of all that, of the five short vowels, only four of them /ɪ, ɛ, ɐ, ɔ/ appear in all positions in Arabic-derived lexicon. In this portion of the lexicon, /ʊ/ occurs only in final position in unstressed syllables, with the single exception of *kull* 'all'. Were it not for the (extensive) Italo-Romance component of the Maltese lexicon, therefore, we can say that the distinction between [ɔ] and [ʊ] would remain allophonic, as it is in Tunis Arabic. As it is, the two sounds should probably be considered phonemically distinct in Maltese. Although minimal pairs are hard to find, possible examples include *punt* 'point' vs. *pont* 'bridge' and *lotto* 'lottery' vs. *luttu* 'mourning'.[7]

Among the long vowels, the presence of /ɛː/ and /ɔː/ phonemes in Maltese is also largely attributable Italo-Romance loans containing these sounds. Although /ē/ and /ō/ occur in certain Tunisian Arabic varieties (**gibson2011**; **herinzammit2017**), these are the result of historical monophthongization of the original *ay and

---

[6]These latter changes are themselves, however, arguably contact-induced – see §3.1.2.

[7]Our thanks to Michael Spagnol for suggesting these examples.

\*aw diphthongs. The Maltese reflexes of these sounds remain diphthongs, as in *sejf* /sɛɪf/ 'sword' and *lewn* /lɛʊn/ 'colour'. Other than in cases of compensatory lengthening in items where the consonants represented by <għ> and <h> have been lost (see §3.1.2), /ɛː/ and /ɔː/ only occur in the non-Arabic component of the Maltese lexicon, as in *żero* /ˈzɛːrɔ/ 'zero' and *froġa* /ˈfrɔːdʒɐ/ 'omelette'.

To these three contact-induced monophthongal innovations we can add one new contact-induced diphthong: /ɔɪ/. Mifsud (**mifsud2011**) points out that this occurs only in non-Arabic lexical items (e.g. *vojt* /vɔɪt/ 'empty space') in standard Maltese.

In summary, then, the majority of innovative vowel phonemes in Maltese are not the direct result of transfer, but the three new monophthongal phonemes whose emergence is (at least partially) contact-induced, combine to create a nearly symmetrical system in which all five short vowel phonemes have a long counterpart.

## 3.3 Intonation

The study of intonation in Maltese, as in most non-Indo-European languages, remains in its infancy (cf. Hellmuth, this volume). Impressionistically speaking, the tunes that can be heard in Maltese (and Maltese English) speech are highly distinctive, and often quite unlike those of the Mediterranean Arabic dialects. Several studies have demonstrated that intonation patterns are highly susceptible to transfer in language contact situations, especially through imposition by SL-dominant speakers (see the studies of Spanish intonation by **orourke2005**; **gabrielkireva2014**). Interestingly, however, this appears to be less true for the tunes associated with polar interrogatives, at least in the varieties of Spanish described by the aforementioned authors, presumably because of the importance of intonation in establishing interrogative force in the absence of syntactic cues in this language. What data we have on this issue for Maltese fits rather neatly into this larger picture. According to **vella2003**, the intonational patterns of Maltese late-focus declaratives on the one hand, and wh-interrogatives on the other, pattern with Palermo Sicilian and Tuscan Italian respectively, while that of Maltese polar interrogatives more closely resembles counterparts in Arabic dialects.

It seems safe to assume that imposition by native speakers of Italo-Romance varieties is the primary cause of the similarities between Maltese and Italo-Romance intonation, but borrowing by Maltese-dominant bilinguals should not be ruled out as an additional factor.

# 4 Morphology

## 4.1 Nouns and adjectives

### 4.1.1 Inflection

It has been shown (e.g. **gardani2012**; **seifart2017**) that plural affixes are, with case affixes, the most widely transferred inflectional morphemes. Maltese conforms neatly to the general crosslinguistic picture: it has acquired plural morphemes from Sicilian and English and little in the way of other inflectional morphology (but see §4.2).[8]

In addition to a rich array of stem-altering (so-called "broken") plural patterns, most of which also serve as the plurals of at least some items of Italo-Romance or, more rarely, English origin (see **spagnol2011** for details), Maltese has six plural suffixes: *-in, -a, -iet, -ijiet, -i,* and *-s.*[9] Of these, *-in,* and *-iet* are straightforward retentions from Arabic (nevertheless extended to numerous non-Arabic items), *-i* and *-s* are straightforward cases of indirect affix borrowing (in the sense of **seifart2015**), and *-a,* and *-ijiet* arguably involve a subtle interplay of internal and externally caused developments.

The most recently borrowed plural suffix is the English-derived *-s.* This occurs exclusively with bases borrowed from English, and may be considered only partially integrated into monolingual Maltese (to the extent that such a thing exists; see §2.2), in that it often alternates optionally with *-ijiet* in items such as *kejk* 'cake' (pl. *kejkijiet ~ kejks*). There are, however, a number of reasonably frequent items (e.g. *friżer* 'freezer') which appear never to take a plural suffix other than *-s.*

The Sicilian-derived suffix *-i* can mark the plural of a far higher proportion of Maltese nouns than can *-s*, and is demonstrably better integrated into the Maltese inflectional system. In addition to marking the plural of Sicilian-derived nouns which also take *-i* (e.g. *xkupa* 'broom' < Sicilian *scupa* (pl. *scupi*); *fjakk* 'weak' < Sicilian *fiaccu* (pl. *fiacchi*)), it has also been extended to: Italian-derived nouns (including those with a plural in *-e* in Italian, e.g. *statwa* 'statue' < Italian *statua* (pl. *statue*)); nouns from other Romance languages (e.g. *pitrava* 'beetroot' < French *betterave* with ∅-plural (orthographic *-s*)); English-derived nouns (e.g. *jard* 'yard (unit of distance)'), and even a few Arabic-derived nouns (e.g. *saff* 'layer' < *ṣaff*

---

[8]One should note also, however, the appearance in a couple of items of a singulative suffix *-u*, apparently borrowed from Sicilian. **borg1994** cites *wiżż-u* 'geese-sɢᴠ', *dud-u* 'worms-sɢᴠ', and *ful-u* 'beans-sɢᴠ'.

[9]There are also one or two examples of zero plurals, e.g. *martri* 'martyr(s)'.

'row', *samm* 'very hard' < *ʔaṣamm* 'deaf, hard').

Arabic and Sicilian coincidentally have an identical less frequently used plural (or collective) suffix *-a*, as in Arabic *mārra* 'passers-by' (singular *mārr*) and Sicilian *libbra* 'books' (singular *libbru*). A plural suffix of this form also occurs in Maltese, with nouns of both Arabic and Italo-Romance origin (e.g. *kittieba* 'writers' < Arabic *kattāb*; *nutara* 'notaries' < Italian *notaro*). Evidence that this is perceived and treated as a single morpheme rather than two homophonous items comes from the fact that the restriction of this suffix to groups of people in Arabic applies also to the Italo-Romance part of the Maltese lexicon (**mifsud2011**).

A curious feature of Maltese plural morphology from a comparative Arabic perspective is the very frequent suffix *-ijiet* (*-jiet* after certain vowel-final stems), as in *postijiet* 'places' (singular *post*) and *ommijiet* 'mothers' (singular *omm*). While clearly based on the Arabic-derived suffix *-iet* (< Arabic *-āt*, with characteristic Maltese *imāla*), the provenance of the initial *-ij-* is not obvious. Mifsud (**mifsud2011**) plausibly suggests that *-ijiet* as a whole is "derived from the plural of verbal nouns with a weak final radical, like *tiġrijiet* 'races', *tiswijiet* 'repairs' ", but **geary2017** makes a strong case that the large influx into Maltese of Italo-Romance nouns whose singulars ended in *-i* (e.g. *affari* 'affair, matter' < Sicilian *affari* or Italian *affare*) was instrumental in the emergence of this morpheme. On this account Maltese speakers originally pluralized such words with *-iet*, with glide-insertion an automatic phonological consequence of the juncture of a vowel-final stem and a vowel-initial suffix. Later, according to Geary, the whole string *-ijiet* was reanalysed as constituting the marker of plurality, and this new plural suffix was extended to consonant-final stems, including Arabic-derived items of basic vocabulary such as *omm* 'mother' and *art* 'land'.[10]

### 4.1.2 Derivation

Maltese displays a rich array of derivational suffixes borrowed (presumably initially as part of polymorphemic lexical items) from Italo-Romance. A definitive list of these has not been provided to date, but **saadeforthcoming** offers a detailed

---

[10]Geary's contact-induced scenario for the emergence of this suffix may not be the whole story, however. Evidence on this point comes from Arabic loanwords in Siwi Berber. Souag (**souag2013**) lists a number of examples of Arabic-origin nouns whose plural is formed by adding a suffix *-iyyat* (e.g. *ṣḥilfa* 'turtle', pl. *ṣḥilfiyyat*), despite the fact that both Classical Arabic and present-day Egyptian Arabic lack plurals of this type. Siwi must therefore have borrowed these items and their pluralization strategy from some early form of (eastern) Maghrebi Arabic, suggesting that the presence in Maltese of the *-ijiet* suffix is, at least to some extent, an Arabic-internal development that predates the large-scale borrowing of Italo-Romance nouns into Maltese.

typology of such items, of which we present a simplified version here, drawing also on examples from **brincatmifsud2015**, and focusing just on the nominal, adjectival and adverbial domains (see §4.2.2 for borrowed participial morphology).

First of all, there are at least twenty suffixes, such as the nominalizer *-zzjoni*, which, though relatively frequent, only occur in items clearly borrowed wholesale from Italo-Romance (e.g. *dikjarazzjoni* 'declaration' < Italian *dichiarazione*) or in coinages which, in a process that is relatively common in Maltese, represent borrowings from English that are adapted to fit the phonology and morphology of Romance-influenced Maltese, as in *esplojtazzjoni* 'exploitation' (cf. **gattfabri2018**). Given this restriction, there must be some doubt as to whether one can regard the suffixes themselves as borrowed, or only the polymorphemic items in which they occur.

Secondly, there are a number of borrowed suffixes which are sufficiently well integrated that they can attach to Arabic-derived bases. Examples include:

> *-ata*, e.g. *xemxata* 'sunstroke' (*xemx* 'sun')
>
> *-ezza*, e.g. *mqarebezza* 'naughtiness' (*mqareb* 'naughty')
>
> *-un* (< Sicilian *-uni*, Italian *-one*), e.g. *ħmarun* 'great fool' (*ħmar* 'donkey')

Finally, there is at least one borrowed suffix *-tura*, which forms a single-instance verbal noun, whose integration can be seen from the fact that it attaches to productively to English bases, as in *ċekkjatura* 'an instance of checking' or *weldjatura* 'an instance of welding'.

## 4.2 Verbs

### 4.2.1 Loaned verbs

Maltese has borrowed a large number of verbs from Sicilian and Italian, and more recently a smaller number from English. The chief interest in these borrowings lies in the way in which they have been integrated into the Maltese inflectional and derivational verbal paradigms. An in-depth study of this phenomenon was provided by **mifsudloanverbs**, who distinguished the following four types of loaned verbs:

> Type A: Full integration into Semitic Maltese sound verbs
>
> Type B: Full integration into Semitic Maltese weak-final verbs

Type C: Undigested Romance stems with a weak-final conjugation

Type D: Undigested English stems

Mifsud (**mifsudloanverbs**) points out that most (perhaps all) Type A verbs are so-called "second generation" loans, whereby a nominal or adjectival form has been borrowed, a root extracted from it, and a verb formed on this root, as in *pitter* 'to paint' – a denominal derivation from *pittur* 'painter', borrowed from Sicilian *pitturi* (and supported by Italian *pittore*). Such items do not, therefore, represent genuine cases of transfer of verbs, and are reminiscent of similar coinages in other Arabic varieties (e.g. *fabrak* 'to fabricate'). In Arabic as in Maltese, such items are overwhelmingly restricted to the denominal verbal stems II and V of triliteral roots and I and II of quadriteral roots (CVCCVC and tCVCCVC).

In contrast to Type A, Mifsud's Types B and C are genuine cases of loaned verbs. Mifsud (**mifsudloanverbs**) shows that the imperative (rather than the homophonous 3SG present, or any other verb forms) was the most likely base form of the Romance models on which the Maltese loaned verbs were created.[11] In both Italian and Sicilian all verbs in the imperative end in either *-i* or *-a.* As it happens, Maltese weak-final verbs (in which the final radical element is a vowel rather than a consonant) also all end in either /ɪ/ or /ɐ/ in the imperfect and imperative singular, depending on which of the two weak-final conjugation classes they fall into. This coincidence resulted in borrowed Romance verbs being integrated into one of the these two weak-final classes, as in *kanta* 'he sang', *jkanta* 'he sings' (< Sicilian/Italian imperative *canta*); and *serva* 'he served', *jservi* 'he serves' (< Sicilian/Italian imperative *servi*).

The difference between verbs of Types B and C is that the former are analysed as having root-and-pattern morphology, with a triliteral or quadriliteral root, whereas Type C are borrowed as a concatenative stem without a root. This can be seen from the fact that Type B verbs can give rise to new verbs with the same root in other verbal stems, as in *kompla* 'to continue', *tkompla* 'to be continued' (< Sicilian *cumpliri* 'to finish'), whereas Type C verbs cannot.

Another difference between Types B and C is that no Type C verb begins with a single (ungeminated) consonant, whereas most Type B verbs do. In fact, apart from certain well-defined exceptions (see **mifsudloanverbs**), all Type C verbs begin with a geminate consonant, as in *ffolla* 'to crowd' < Italian *affollare*. What exactly was the combination of historical factors that gave rise to this synchronic state of affairs is a complex matter (see **mifsudloanverbs** for discussion), but

---

[11]This parallels the situation in Arabic-based pidgins and creoles, for which **versteegh2014** shows that verbs generally appear to derive from infinitives in the lexifier varieties.

the key point to note is that at least some of the instances of initial gemination in Type C verbs are apparently not attributable to phonological properties of the source item (e.g. *pprova* 'to try' < Italian *provare*). It seems that speakers of Maltese came to feel that all loan verbs must have an initial geminate consonant, whether or not this was actually true of the item being borrowed.

This state of affairs manifests itself rather spectacularly in more recent borrowings from English (Type D verbs), in which initial consonants are duly geminated (despite this never being the case in the English source items), but which also fall into the conjugation class of weak-final verbs, as in *ddawnlowdja* 'to download'. What underlies this treatment of loans from English seems to be a type of reanalysis, which we can sketch as follows. In the initial stage, verbs without roots (not necessarily identifiable to speakers as loans from Italo-Romance) are analysed as falling into the weak-final conjugation class because they have a stem-final vowel. But since all verbs without roots (at this pre-English stage) have a stem-final vowel, it is possible to view the lack of a root, not the presence of a stem-final vowel, as the reason that loan verbs obligatorily fall into the weak-final conjugation class; and it seems that speakers indeed made this reanalysis. In a parallel development, initial consonant gemination also came to be seen an obligatory feature of the class of verbs lacking a root. As a result of these developments, when a verb is borrowed from English, because it lacks a root its initial consonant is geminated and it is conjugated as a weak-final verb, regardless of whether it has a stem-final vowel.[12]

### 4.2.2 Participles

Unsurprisingly, one of the additional ways in which Type A verbs differ from the remaining three classes of loaned verbs is the formation of passive participles: in Type A verbs, passive participles are formed in accordance with the Semitic pattern for the respective derived stem, e.g. *pejjep* 'to smoke' (stem II, from Italian *pipa* 'pipe') produces *mpejjep* 'smoked' (**mifsudloanverbs**). In contrast, some Type B verbs allow for the formation of a passive participle using Romance suffixes (**mifsudloanverbs**), and this is the sole option for Type C and even Type D verbs: for Type C verbs, the choice of the actual suffix depends on the original form of the verb and, in some cases, the path of borrowing (see below). For

---

[12]In addition, virtually all Type D verbs insert a palatal glide between the borrowed stem and the added weak-final vowel, as in *pparkja* 'to park'. Similarly to the initial gemination and weak-final inflection of Type D verbs, this glide insertion must be the result of analogical extension from numerous glide-final borrowed Romance verbs, e.g. *rdoppja* 'to double' < Italian *raddoppiare*. See **mifsudloanverbs** for a detailed discussion.

Type D verbs borrowed from English, the suffix *-at* is the only productive way to form a passive participle (e.g. *inxurjat* 'insured') with *spellut* 'spelled' as the only exception (**mifsudloanverbs**).

And finally, there are two distinct classes of Type B and C verbs which can each derive two passive participles. In the first class, one participle is derived from the weak (regular) form root and the other derived from the strong one, e.g. *konfondut* 'confused' vs. *konfuż* (**mifsudloanverbs**). In the second class, one participle is derived using the Sicilian suffix *-ut*, the other using the Standard Italian suffix *-it*, e.g. *preferut* 'preferred' vs. *preferit* (**mifsudloanverbs**). The reason for these doublets is largely sociolinguistic: the variability of the first class echoes a similar situation in Italian dialects (**mifsudloanverbs**); that of the second class reflects a situation whereby the loaned verb effectively has two sources, spoken Sicilian and Standard (Tuscan) Italian.

## 5 Syntax

### 5.1 Phrase syntax

#### 5.1.1 Word order

The expansion of Maltese lexicon with words borrowed from Sicilian and Italian had a profound effect on the syntax of Maltese. The primary example of this is word order within the noun phrase, involving the order of adjectives and their heads. In Arabic, adjectives (with the exception of comparatives, superlatives and a number of specific cases) follow their heads. This is largely true of Italian adjectives as well, with the exception of a small subclass some grammars term "specificational adjectives" (e.g. **italian2007**), such as *stesso* 'same' and *certo* 'certain', which precede their head. Such adjectives borrowed into Maltese retained their syntactic properties, as with the pre-nominal *ċertu* (< Sic. *certu*) in (3).

(3) Kien      bniedem ta' ċerta    personalità.
    be.PRF.3SG.M person   GEN certain.F personality.
    'He was a person with a certain personality.'
    [BCv3: it-torca.8685]

In Italian, specificational adjectives to a large extent overlap with a class of adjectives that perform double duty as quantifiers (or perhaps determiners) and vary their position according to their respective roles: Adj–N for quantifiers, N–Adj for adjectives. One could argue that it is in the former function that they were borrowed into Maltese and thus should be considered quantifiers or determiners

rather than adjectives, especially in light of the fact that they are (for the most part) in complementary distribution with the definite article, as determiners and quantifiers are. Determiners and quantifiers in Maltese precede their heads (as with the definite article *il-*, *kull* 'all', *xi* 'some' etc.).

There are three arguments against such an account: first, borrowed pre-nominal specificational adjectives actually fall into two classes, where members of the first, such as *ċertu* 'certain', *diversi* 'diverse' (< It. *diverso*) or *varju* 'various' (< Sic. *varju*), do not (for the most part) allow the definite article. In contrast, words in the second class such as *stess* 'same' (< It. *stesso*) or *uniku* 'unique' (< Sic. *uniku*) predominantly co-occur with the definite article when pre-nominal. The same, incidentally, is true of the etymologically Arabic pre-nominal quantifier *ebda* 'no, none'.

Secondly, there are morphological considerations: pre-nominal specificational adjectives of both types mark gender and/or number (*varju* for the first, *uniku* for the second) like Maltese adjectives do; Maltese determiners and quantifiers do not inflect for either gender or number.[13]

The final argument against considering borrowed pre-nominal specificational adjectives as being borrowed into the slot for determiners involves ordinal numerals. In Italian, these also fall into the subclass of prenominal specificational adjectives (**italian2007**) and thus precede their head. The same is invariably true of Maltese ordinal numerals, as with *ewwel* in (4).

(4) Wara l-ewwel sena
after DEF-first year
'After the first year'
[BCv3: l-orizzont.64586]

In North African Arabic, ordinal numerals can either precede or follow their heads, but when they precede them, they never take the definite article, even when the noun phrase is semantically definite (see e.g. **veronika2014** for Tunisian Arabic). In contrast, Maltese never allows its ordinal numerals to follow their heads, and the definite article is obligatory.

All these arguments, including the comparison with related Arabic varieties, suggest that the pre-nominal position of some adjectives and ordinal numerals in Maltese is due to transfer under RL-agentivity from Italian.

---

[13]With the exception of the very specific category of demonstrative pronouns where gender and number are marked not by affixes, but rather a form of suppletion.

### 5.1.2 The analytical passive

As with adjectives (§5.1.1), lexical borrowings from Italo-Romance have also had a significant impact on the syntax of Maltese verbs. One of the most conspicuous consequences of this development involves the passive voice: as Romance-origin verbs cannot generally form one of the passive derived verbal stems (but see §4.2.1), they brought with them their Romance syntax and thus a new type of passive construction arose in Maltese – the analytical passive.

In Maltese, there are two types of analytical passive construction containing a passive participle: the so-called "dynamic passive" (**borgazzopardimaltese**, **vanhove1993**), which combines passive participles with the passive auxiliary *ġie* 'to come'; and the so-called "stative passive" (**borgazzopardimaltese**, **vanhove1993**), which has the same structure as copular clauses (see §5.2.3), the only difference being that stative passive constructions can feature an agentive NP introduced by the preposition *minn* 'from' (see **bulbul2018** for a detailed analysis).

The stative passive can be viewed as an extension of the structurally identical construction which is sporadically attested already in Classical Arabic (**ullmann1989**), but becomes quite prominent in Christian Arabic documents at least as early as 10th century, where, incidentally, it gained prominence under influence from Aramaic and Greek (**blau1967**).

The dynamic passive (5), on the other hand, is a straightforward calque on either Italian or Sicilian, where a construction featuring a verb semantically equivalent to *ġie* – *venire* in Italian – combines with a past participle (cf. Manfredi, this volume).

(5) Kif diġà  għedt,     ġie            ppreżentat            il-kuntratt.
    as  already say.PRF.1SG come.PRF.3SG.M present.PTCP.PASS.M DEF-contract.

    'As I already said, the contract was presented.'

    [MUDTv1: 30_01P05]

While the dynamic passive must have originally functioned to fill a hole in the verbal system of Maltese by providing a way to passivize Romance verbs, it has meanwhile spread to include native verbs as well, as with *ta* 'to give' (< root ʕ-t-y) in (6).

(6) It-tagħrif         ġie            mogħti        mill-Ministru
    DEF-information come.PRF.3SG.M give.PTCP.PASS from.DEF-minister
    Konrad Mizzi.
    Konrad Mizzi.

    'The information was given by Minister Konrad Mizzi.'

[BCv3: inewsmalta-ott.29.2013.1257-11045]

### 5.1.3 Modality

Another clear-cut example of grammatical calquing comes from the domain of modality and involves the pseudoverb *għand-*. In Maltese, its primary function is that of a possessive (7), as is the case with its cognates *ʕind-/ʕand-* in many Arabic varieties.

(7)  M' għandi  xejn    kontri-hom.
     NEG have.1SG nothing against-OBL.3PL

     'I have nothing against them.'

     [MUDTv1: 22_02J03]

In addition to this, however, the Maltese *għand-* has also taken on a function as a deontic modal of weak obligation 'should, ought' taking verbal complement, as in (8).[14]

(8)  Naqbel        li    għandhom jivvutaw       aktar nies.
     agree.IMPF.1SG COMP have.3PL  vote.IMPF.3PL.M more people.

     'I agree that more people should vote.'

     [MUDTv1: 22_02J03]

The use of *għand-* in this kind of modal function appears to be unique to Maltese; not even Cypriot Maronite Arabic with its many parallels to Maltese (on which see below) exhibits the same behavior for its cognate *ʕint-* (**borg2004**) and uses a different verb, *salaħ/pkyislaħ* (**borg2004**), as the default deontic modal. The Maltese development must therefore be another calque, since the basic possessive verb of Sicilian, *aviri*, also doubles as a deontic modal, as in (9).

(9)  Sicilian (**piccitto1977**)
     Cci         l'àiu                  a-ddiri   a-tto     patri.
     DAT.3SG.M ACC.3SG.M-have.PRES.1SG to-say.INF DAT-2SG.M father

     'I have to say it to your father.'

---

[14]*għand-* is the only Maltese pseudoverb (and verb) which exhibits a three-way distinction between present (*għand-*), past (*kell-*) and future/habitual (*ikoll-*) forms; all can occur in the modal function.

## 5.2 Sentence syntax

### 5.2.1 Differential object marking

Differential object marking (DOM) is a phenomenon where direct objects are marked according to some combination of semantic and pragmatic properties of the object in question. In Spanish, for example, objects denoting humans (and equivalent entities) are marked by the particle *a*, originally a directional preposition. DOM is a phenomenon attested cross-linguistically (see **khan1984** for Semitic languages), including in varieties of Arabic like Levantine and Iraqi Arabic (**coghill2014** and references therein) and in Andalusian Arabic (**andalusi2013**).

DOM is a well-documented feature of Maltese morphosyntax and largely conforms to the Spanish prototype: in general, both pronominal and nominal direct objects denoting entities high in the "animacy hierarchy" (**borgazzopardimaltese**) take the object marker *lil* (10), which also does double duty as the indirect object marker for all objects. Inanimate objects do not take *lil* (11).

(10)  Min jara      lili    jara        lil Missier-i
      who see.3SG.M.IMPF ACC.1sg see.IMPF.3SG.M ACC father-OBL.1SG
      'Who looks at me, looks at my Father'
      [BCv3: ilgensillum.2011-Mejju-22.8230]

(11)  Min jara          orrizzonti ġodda u   min baħħ.
      who see.IMPF.3SG.M horizon.PL new.PL and who void.
      'Some see new horizons, some see a void.'
      [BCv3: l-emigrant]

**dohla2016** examines DOM in Maltese in some detail and arrives at the conclusion that while there is "a certain predisposition for object marking in general within pan-Arabic grammar" (**dohla2016**), Maltese DOM cannot be ascribed to purely internal developments within Neo-Arabic. A striking feature of the Arabic varieties that exhibit DOM is that they were all in prolonged contact with other languages: Aramaic for Levantine and Iraqi Arabic (and, by extension, for Cypriot Maronite Arabic, cf. **borg2004**), Romance for Andalusian Arabic and Maltese. In the case of Maltese, said Romance variety is Sicilian, where the object marker *a* performs the same double duty as the Maltese *lil*, and DOM in both languages shows a number of remarkable similarities: in both Sicilian and Maltese, DOM is primarily triggered "by humanness along with definiteness/referentiality" (**iemmolo2010** in reference to Sicilian), it is obligatory with personal pronouns, but optional with plural "kinship terms and human common nouns" and disallowed with

"(in)animate and indefinite non-specific nouns" (**iemmolo2010** of Sicilian), as exemplified by the non-specific Maltese *nies* "people" in (12).

(12)  Min irid          jara          nies jgħixu        hekk?
      who want.IMPF.3SG.M see.IMPF.3SG.M people live.IMPF.3PL.M thus?
      'Who wants to see people live like that?'
      [BCv3: l-orizzont.41390]

In Maltese DOM, then, we have an instance of what Manfredi (this volume) labels "calquing of polyfunctionality of grammatical items inducing syntactic change": Maltese acquired a rule of DOM as a result of the indirect object marker *lil* inheriting the dual function of its Sicilian equivalent *a*. It is clear that this is a contact-induced change. But since with this and the similar changes discussed below there is no transfer of lexical matter, it seems impossible at present to judge whether they are the result of borrowing or imposition, or whether they were actuated by speakers for whom neither the RL or the SL were dominant, in the process that Lucas (**lucas2015**) calls "convergence".

### 5.2.2 Clitic doubling (proper)

The existence of various reduplicative phenomena associated with direct and indirect clitic pronouns in Maltese has been noted at least since **sutcliffe**, who identifies what classical tradition refers to as *nominativus pendens*. This analysis has been elaborated on by **fabri1993**, **borgazzopardimaltese** and **fabriborgtopicfocus**, primarily in the context of pragmatically determined constituent order variation, especially topicalization. Building on these works and the analysis of Maltese clitics by **camilleri2011**, **bulbul2014** notes that in addition to these phenomena, which in one way or another entail dislocation, there exists in Maltese another related phenomenon, where lexical objects and clitic pronouns co-occur, but without the dislocation of the lexical object. This phenomenon, termed Clitic Doubling Proper to distinguish it from similar constructions (see **krapovacinque2008** for a detailed analysis), involves the co-occurrence of a lexical object and the clitic with the object in situ, which in Maltese is after the verb (see **bulbul2018**). Maltese Clitic Doubling Proper occurs with both direct (13) and indirect objects (14).

(13)  Ftit nies    jafu-ha                     l-istorja         marbuta  ma' dan
      few people know.IMPF.3PL.M-3SG.F DEF-history.SG.F connected with this
      il-proġett   tant sabiħ.
      DEF-project  such beautiful.

      'Few people know the history connected with such a beautiful project.'

      [BCv3: l-orizzont.36758]

(14)  Hekk qed  ngħidu-lhom          lil  dawn in-nies      f'
      thus  PROG say.IMPF.1PL-DAT.3PL DAT this.PL DEF-people in
      pajjiż-na.
      country-OBL.1PL

      'This is what we say to these people in our country.'

      [BCv3: 20020313_714d_par]

Unlike various types of dislocation with resumptive clitic pronouns which are quite common in European languages (see e.g. **decat2010**), Clitic Doubling Proper is a much rarer phenomenon; in Europe, it is largely confined to the Balkan *Sprachbund* (**friedman2008**) and some Romance languages outside of the Balkans, like Spanish (**zagona2002**) and varieties of Italian (**russi2008**). The phenomenon is also attested in Semitic languages (**khan1984**), including Arabic, where it was studied in detail by **souagcliticdoubling**. Comparing Clitic Doubling Proper in various varieties of Arabic including Maltese, Souag (**souagcliticdoubling**) notes parallels between Maltese and some varieties of Algerian Arabic, especially in regard to the doubling of indirect objects. Ultimately, however, he arrives at the conclusion that Maltese Clitic Doubling Proper "has little in common with any other Arabic variety examined, but closely resembles that found in Sicilian" (**souagcliticdoubling**). This suggests that here too we have a contact-induced change, this time of the sort that Manfredi (this volume) labels "narrow syntactic calquing", that is, without any accompanying calque of lexical items.

### 5.2.3 Copular constructions

In Maltese, there are four types of copular clauses (**borgazzopardimaltese**):[15]

    Type 1: No copula

---

[15]In addition to these, **borg1987** and **borgspagnol2015** also describe the copular function of the verb *jinsab* 'to be found'. This being a finite verb, both **borgazzopardimaltese** and **bulbul2018** exclude this type of clause, as well as similar ones, such as those featuring the verb *sar* 'to become', from the category of copular clauses.

Type 2: The verb *kien* as the copula

Type 3: Personal pronoun as the copula

Type 4: Present participle *qiegħed* as the copula

Type 1 describes what traditional grammars of Semitic languages refer to as nominal sentences; type 2 then contains their counterparts with a past reference. Types 3 and 4, while not without parallel in other varieties Arabic,[16] feature much more prominently in Maltese. This is especially true of type 3 copular clauses, which involve the use of a personal pronoun as the copula (15).

(15)  Din   hi   omm-ok.
       this.f 3sg.f mother-obl.2sg.
       'This is your mother.'
       [BCv3: 2010 Immanuel Mifsud - Fl-Isem tal-Missier (U tal-Iben)]

While such copular constructions have been described for Egyptian (**eid1983**) and Syrian (**berlinches2016**) Arabic, they are highly uncommon in North Africa.[17] It is therefore even more significant that in Maltese they have become the dominant form of copular construction in the present time-frame: in MUDTv1, for example, 110 non-negative copular clauses are of Type 1; 181 are Type 3. In this, Maltese type 3 copular clauses are comparable to equivalent copular constructions in Anatolian Arabic (see **lahdo2009** for Tillo Arabic and the references therein, as well as Akkuş, this volume), Andalusian Arabic (**andalusi2013**), and especially Cypriot Maronite Arabic (**borg1985**; Walter, this volume), where they are but one piece of evidence linking Cypriot Maronite Arabic to *qəltu* dialects (**borg2004**). The conclusion to be drawn here is the same as for DOM and Clitic Doubling Proper above: it is no coincidence that these copular constructions are in wide use and the copular construction of choice only in varieties of Arabic which have been under contact influence from languages with a mandatory copula – Turkish for Anatolian Arabic, Spanish for Andalusian Arabic, Greek for Cypriot Maronite Arabic, and Italian for Maltese. Whether the origin of such constructions can be traced to a feature in (one of) these dialects' Old Arabic ancestors, or whether they came about through parallel development, contact undoubtedly triggered the widespread adoption of such constructions in these varieties of Arabic.

---

[16]See the analysis of Type 4 copulas in **camillerisadler2018**.

[17]On the other hand, Lameen Souag informs us that such constructions do occur in the Algerian Arabic of Dellys.

# 6 Lexicon

## 6.1 Major sources

That Maltese contains large numbers of loanwords from Romance and English is a fact immediately obvious to even the most casual observer. Over the years, there have been a number of attempts to quantify the influence of other languages on Maltese by providing a classification of lexemes by their origin. The earliest, **fenech1978**, compiled such statistics for journalistic Maltese, but also provided a comparison to literary and spoken Maltese (albeit using a very small data sample). Brincat analyzed the etymological composition of entries in Aquilina's dictionary, first examining the origin of 34,968 out of all 39,149 headwords (**brincat1996**) and then applying the same analysis to the entire list (**brincat2011**); **mifsudborg1997** did the same with the vocabulary contained in an introductory textbook of Maltese as a foreign language. In 2006, **bovingdondalli2006** analyzed the etymology of lexical items in a 1000-word sample obtained from a corpus of Maltese and most recently, **comriespagnol2016** did the same on a list of 1500 "lexical meanings" within the framework of the *Loanwords in the world's languages* project (**loanwords2009**). Figure 1 summarizes all these findings.
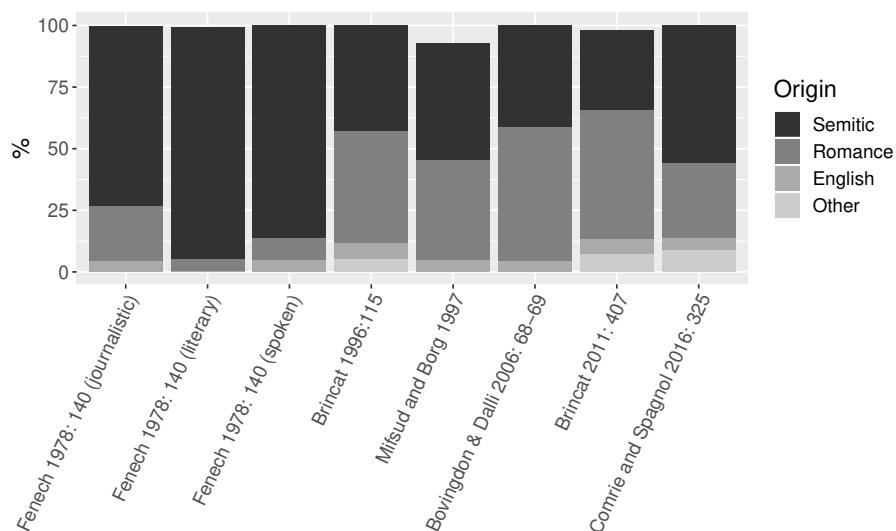


Figure 1: A summary of previous studies of the composition of the Maltese lexicon

The primary explanation for the sharp differences between these analyses is methodology: while **fenech1978** analyzes entire texts and thus counts tokens, **brincat1996** (including its updated version in **brincat2011**) and **bovingdondalli2006** analyze lists of unique words, i.e. types. The later is also true of **mifsudborg1997** and **comriespagnol2016**, except where **brincat1996** uses dictionary data and **bovingdondalli2006** corpus data, **mifsudborg1997** employ a list of lexical items with high frequency of use in daily communication and **comriespagnol2016** base their analysis on a list compiled for the purposes of cross-linguistic comparison. The high ratio of words of Semitic origin in token-based analyses is thus due to the prevalence of function words, which are overwhelmingly Arabic. The type-based analyses then provide a somewhat more accurate picture of the lexicon as a whole, even though they are not without their problems. Chief among these is the issue of what exactly counts as type, especially with regard to productive derivational affixes, e.g. whether all the words with the prefix *anti-* count as distinct types or not.

In addition to general analyses, **bovingdondalli2006** and **comriespagnol2016** also provided breakdowns for individual parts of speech. Unfortunately, these analyses are not comparable, as each has a different focus: **bovingdondalli2006** are interested in the composition of each etymological stock by word class (Table 3).

Table 3: Source language component of Maltese by word class (**bovingdondalli2006**).

| Origin | Function words | Verbs | Adjectives | Nouns | Adverbs | Pronouns |
|---|---:|---:|---:|---:|---:|---:|
| Semitic | 3% | 70% | 2% | 21% | 2% | 2% |
| Romance | 0% | 38% | 11% | 48% | 3% | 0% |
| English | 0% | 29% | 8% | 63% | 0% | 0% |

In contrast, Comrie & Spagnol **comriespagnol2016**) focus on the composition of individual word classes by their origin (Table 4).[18]

---

[18]The details of Comrie & Spagnol's (**comriespagnol2016**) methodology mean that loans in their dataset come from Romance and English but not from any other languages. The category we label "Misc." in Tables 4 and 5 encompasses those meanings in the *Loanwords in the world's languages* 1500-item set which have no corresponding single-word Maltese lexical item, and those where the etymology is at present unknown, or where the item in question is an innovative Maltese-internal coinage.

Table 4: Word class composition by source language (**comriespagnol2016**)

| Word class | Arabic | Romance | English | Misc. |
|---|---|---|---|---|
| Function words | 84.7% | 6.2% | 0% | 9.1% |
| Verbs | 75.3% | 14.1% | 1.3% | 9.2% |
| Adjectives | 65.2% | 28.5% | 0.3% | 6.0% |
| Nouns | 44.7% | 39.6% | 7.2% | 8.6% |

Comrie & Spagnol (**comriespagnol2016**) also provide a breakdown of their data by semantic field, permitting a comparison of the domains in which Romance versus English loans are more or less prominent. A number of generalizations can be made here (see Table 5 for a summary), though ultimately they all follow naturally from the fact that contact with English was more recent, and less intensive, than contact with Sicilian and Italian.

Table 5: Composition of semantic fields by source language (**comriespagnol2016**)

| Semantic field | Arabic | Romance | English | Misc. |
|---|---|---|---|---|
| Modern world | 3.0% | 65.3% | 22.8% | 9.0% |
| Animals | 47.8% | 29.1% | 13.9% | 9.1% |
| Clothing and grooming | 38.7% | 47.2% | 10.4% | 3.8% |
| Warfare and hunting | 28.8% | 65.0% | 2.5% | 3.8% |
| Law | 36.0% | 50.0% | 0.0% | 14.0% |
| Social and political relations | 48.4% | 48.4% | 0.0% | 3.2% |

Unsurprisingly, English is best represented in the category of items relating to the modern world, but even here Romance dominates. Examples include English-derived *televixin* 'television' and Italian-derived *kafè* 'coffee'.

The domain of animals divides rather neatly as follows. Common animals (especially land animals) of the Mediterranean area are largely Arabic-derived (e.g. *fenek* 'rabbit' < Maghrebi Arabic *fanak* 'fennec fox'), while well-known non-indigenous animals are largely Romance-derived (e.g. *ljunfant* 'elephant' < Sicilian *liufanti*, the additional /n/ perhaps the result of influence from *ljun* 'lion'). More exotic animals, if there is a corresponding Maltese item at all, derive from English (e.g. *tapir* 'tapir'). Clothing and grooming presents a similar picture,

with Arabic-derived *suf* 'wool', Sicilian-derived *ngwanta* 'glove', and English-derived *fer* 'fur', as does warfare and hunting, with Arabic-derived *sejf* 'sword', Sicilian-derived *xkubetta* 'gun', and English-derived *senter* 'shotgun' (< *centre-breech-loading shotgun*).

The total lack of English loans in the domains of law and social and political relations, at least in Comrie and Spagnol's sample, is remarkable, given the extent to which the English language dominated public life in Malta in the twentieth century. A generalization that underlies this finding is that while English influence is strongest in the spheres of commerce, consumerism and, especially in the 21st century, popular culture (e.g. *vawċer* 'voucher', *ċċettja* 'to chat'),[19] at least as far as Maltese lexicon is concerned, it has not supplanted Italian in the domains of high culture and the affairs of state (e.g. *gvern* 'government' < Italian *governo*, *poeżija* 'poem' < Italian *poesia*).

## 6.2 Minor sources

Considering its location and the nature of population movements in the Mediterranean, it is hardly surprising that the Maltese lexicon also contains borrowings from languages other than Sicilian, Italian and English. The most obvious of these are borrowings from other Romance languages. First among them, as in other European languages, stands Latin, which provided a large chunk of Maltese scientific and technical vocabulary, whether as terminology (e.g. *ego*, *rektum* or *sukkursu* 'underground water'), biological nomenclature (*fagu* 'European beech, *Fagus sylvatica*', *mirla* 'brown wrasse, *Labrus merula*') or set phrases and expressions (*ex cathedra*, *ibidem*). Curiously for a Catholic country, Latin is the source of very little religious vocabulary in Maltese; in this area, Maltese continues to rely almost exclusively on words of Arabic origin. Those Latin words related to religious matters employed in modern Maltese therefore typically refer to minutiae of Catholic Church rituals and procedures, such as *ekseat* 'a bishop's permission for a priest to leave the diocese' (< *exeat*) or *indult* 'a Pope's authorization to perform an act otherwise not allowed by canon law'. Of the few Latin terms related to religion still in common use, *nobis* stands out as a rather curious lexical item: in Maltese, it is used as a (post-nominal) modifier indicating intensity or size, as

---

[19]Until at least 1991, when the Maltese government opened up television broadcasting rights to more than just the single state broadcaster TVM, Italian television stations, whose broadcasts from Sicily could be received in Malta, were very widely watched, and there was consequently considerable Italian influence on Maltese popular culture (**sammut2007**). This influence has waned considerably at the expense of English and American culture since the advent of broadcast pluralism in Malta, and especially with the rise of cable television and online video streaming.

in *tkaxkira nobis* 'a sound thrashing' or *tindifa nobis* 'a thorough cleaning'.

Before the Order of Saint John gained control of Malta, the islands were for more than two centuries a part (whether officially or not) of the Crown of Aragon. As such, one would expect that speakers of Maltese during that era found themselves exposed the languages of the Crown like Catalan, Spanish and Occitan, that and this was then reflected in the Maltese lexicon. In truth, however, there are only a few Maltese words that can clearly be traced to Ibero-Romance. **catalan2017** identify a number of lexical items with Catalan and/or Occitan origins, but note that many of them can also be found in Sicilian, which in most cases can be clearly determined as the origin of the loan. On the other hand, there are Maltese words of obviously Romance origin whose current shape cannot be easily explained by any of the processes by which Sicilian or Italian words were made to conform to Maltese phonology, and where the Catalan or Occitan origin postulated by **catalan2017** may offer a better explanation than that of "local formation" resorted to by previous works. These may include: *boxxla* 'compass' < Catalan *búixola* vs. Italian *bussola*; *frixa* 'pancreas' < Catalan *freixura* 'entrails' and even the very frequent *żgur* 'certain', which, due to its phonology, especially the /g/ (see §3.1.1), points to an origin in Catalan *segur* or Spanish *seguro*, rather than to Italian or Sicilian which, both feature a /k/ in its place. These and other lexical items, onomastics (see **catalan2017**) and even usage (such as the ubiquitous Maltese swear word *l-ostja*, literally 'the host, sacramental bread', which is very atypical for Italian or Sicilian, but has a counterpart in the Spanish *la hostia*) suggest some influence of Ibero-Romance on Maltese which is yet to be thoroughly researched.

The much shorter French occupation of the Maltese islands left very little linguistic trace, and so it is internationalisms in the semantic field of culture (*bonton* 'high society', *etikett* 'etiquette'), fashion (*manikin* 'manequin') and the culinary arts (*fundan* 'fondant', *ragu* 'ragout') where French borrowings in Maltese can be found. The few notable exceptions include *berġa* (< *auberge*), the term used for the residences of langues (chapters) of the Order of Saint John. The most prominent of these palaces, *Berġa ta' Kastilja*, now houses the office of the Prime Minister of Malta, for which the term *Berġa* is often used metonymically. The other two Maltese words of French origin still in frequent daily use both happen to be connected to transportation: *xufier* (< *chauffeur*) 'driver' and *xarabank* (< *char à bancs*) 'bus'. The latter is particularly interesting due to its pronunciation /ʃɐɾɐˈbɐnk/, which indicates that it was borrowed directly from French and not from English (which would give /ʃɛɾɛˈbɛnk/, as well as for its connection to the French-speaking Maghreb, where the same word was in use; this indicates the

possibility that it was brought from there by Maltese expats.

In addition to Romance languages, post-classical Greek, with its ubiquitous presence all across the Mediterranean (including the neighboring Sicily), could not help but leave a trace on Maltese vocabulary, small though it is. Aquilina (**aquilina1976**) gives *Lapsi* 'Feast of Ascension' (< *ανάληψη*) as the solitary example of a Maltese religious term not inherited from Christian Arabic or borrowed from Romance languages. The other two examples of Greek loanwords involve a completely different sphere: the first one is *ħamallu* 'lewd, vulgar person' < Greek *χαμάλης* (**megaleksikon1958**). This word may ultimately be traceable to Arabic (through Turkish), as evident from its other meaning in Greek, 'porter' (< *ḥammāl*). However, the meaning in which it appears in Maltese is unique to the Greek word, indicating that it was borrowed into Maltese from Greek. The other such term is *vroma* 'complete failure, fiasco' which is quite straightforwardly traceable to the Greek *βρόμα/βρώμα* 'dirt, filth' (**megaleksikon1958**).

With regard to the debates on the origin and history of Maltese, borrowings from other Afro-Asiatic languages have long been at the center of attention of Maltese etymological research. Berber is perhaps the most notorious example here, with a number of items cited as having Berber origins by **colin1957** and **aquilina1976**. Aquilina's list is an expansion of Colin's and thus both feature the same conspicuous items which for the most part involve zoology, such as *fekruna* 'tortoise' (< *fekrun*, **naitzerrad2002**) and *gendus* 'bull' (< *agenduz*, **naitzerrad2002**). Additionally, Aquilina postulates a Berber origin for a number of lexical items where this seems questionable. In some cases the items in question are obviously Arabic loanwords in Berber (as with *bilħaqq* 'by the way', quite transparently from Arabic *b-il-ḥaqq* 'in truth'). In other cases subsequent research has argued against a Berber origin. For example, while Aquilina identifies *żenbil* 'a large carrying basket' as having a Berber origin, **borg2004** notes that it can also be found in the Arabic dialect of Aleppo and Arbil, and traces its ultimate origin to Akkadian through Aramaic. A large group of similarities between Maltese and Berber identified by Aquilina involve "Berber nursery language", containing items like Berber *papa* 'bread' / Maltese *pappa*, Berber *pp-spps* or *ppssi* 'urine' / Maltese *pixxa*, and Berber *kakka/qaqah* / Maltese *kakka* (both having to do with defecation). These forms are actually attested cross-linguistically (**ferguson1964**) at least as far north as Slovak (**ondrackova2010**) and cannot thus be considered loans from Berber. Nevertheless, the fact that there is a Berber lexical component in Maltese is well established, and **souagberber2018** has shown that it may be larger than previously thought (e.g. his case for the Berber etymology of the frequent adjective *ċkejken* 'small').

Finally, in addition to Berber, Maltese also contains a small number of words that can be reasonably traced back to Aramaic. Along with obsolescent lexical items such as *żenbil* given above or *andar* 'threshing floor' (**behnstedt2005**), this small list includes the frequent verb *xandara* 'to broadcast, to spread (news)', otherwise unattested in any other variety of Arabic (**borg1996**). This verb is presumably derived from the common Aramaic root *š-d-r* 'to dispatch, send' with cognates in Mandaic (**manddic1963**), Jewish Babylonian Aramaic (JBA, **sokoloff2002**) and Christian Neo-Aramaic (**khan2008**). The insertion of [n] reflects the dissimilation of the geminated [dd] into [nd] (**lipinski1997**); the same phenomenon involving the original geminated [bb] can also account for *żenbil* (cf. JBA *zabbīlā*, **sokoloff2002**). These borrowings could on the one hand strengthen the case for a Levantine substrate in (if not origin of) Maltese, as **borg1996** insists; on the other hand, some of them can also be found in other North African varieties (**behnstedt2005**).

## Further reading

**krier1976** is a short monograph on the influence of Italo-Romance on Maltese phonology, morphology, syntax, and lexicon.
**mifsudloanverbs** gives an in-depth description of Maltese loaned verbs.
**comriespagnol2016** examine lexical borrowing in Maltese in the context of loanword typology crosslinguistically.
**drewes1994** and **stolz2003** explore the question of whether Maltese is properly labeled a "mixed language".

## Abbreviations

| | |
|---|---|
| ACC | accusative (direct object) marker |
| ADJ | adjective |
| BCv3 | bulbulistan corpus malti v3 |
| COMP | complementizer |
| DAT | dative (indirect object) marker |
| DEF | definite article |
| DOM | differential object marking |
| F | feminine |
| GEN | genitive marker |
| IMPF | imperfect (prefixal conjugation) |
| INF | infinitve |

| It. | Italian |
| L1 | first language |
| L2 | second language |
| M | masculine |
| MUDTv1 | Maltese Universal Dependencies Treebank v1 |
| N | noun |
| NEG | negative (particle) |
| NP | noun phrase |
| OBL | oblique |
| PTCP | participle |
| PASS | passive |
| PL | plural |
| PROG | progressive marker |
| PRF | perfect (suffix conjugation) |
| RL | recipient language |
| SG | singular |
| SGV | singulative |
| SL | source language |
| Sic. | Sicilian |

## Acknowledgments

## Primary sources

Maltese examples above are primarily cited from the general corpus of Maltese *bulbulistan corpus malti v3* (accessible at www.bulbul.sk/bonito2, login: guest, password: Ghilm3), as well as from the *Maltese Universal Dependencies Treebank v1* (accessible at www.bulbul.sk/annis-gui-3.4.4/), both described as to their composition and annotation in **bulbul2018**. Each citation is accompanied by an abbreviation identifying the source (BCv3 and MUDTv1, respectively), as well as the specific document where it can be found.

# Chapter 9

# Contact and calquing

## Stefano Manfredi
CNRS, SeDyL

The notion of calquing refers to the transfer of semantic and syntactic patterns deprived of morphophonological matter. By providing examples of lexical and grammatical calques in a number of Arabic dialects and Arabic-based contact languages, this chapter identifies ways to relate the process of calquing to Van Coetsem's psycholinguistic principle of language dominance.

## 1 Introduction

> APA-like comma is problematic for citation retrieval and conversion
> paper should use smallcaps instead of ALLCAPS
> translations should be in single quotes

In its simplest definition, CALQUING is a type of contact-induced change in which a word or sentence structure is transferred without actual morphemes (**Thomason2001**). Calques are sometimes called loan translations as they typically represent a word-by-word (or morpheme-by-morpheme) translation of a lexeme or a sentence from another language. **Heath1984** labels this process "pattern transfer" and distinguishes it from "matter borrowing" which is instead linked to the integration of morphophonological material. **Ross2007**, for his part, points out that that calquing can also produce important grammatical changes, and he considers it a necessary precondition for contact-induced morphosyntactic restructuring (what Ross calls "metatypy").

Broadly speaking, we can distinguish two types of calquing: lexical calquing, which entails the transfer of semantic properties of lexical items, and grammatical calquing, which instead implies the transfer of the functional properties of morphemes and syntactic constructions. Using Ross's words (**Ross2007**), lexical

calquing consists of remodelling lexical "ways of saying things", whereas grammatical calquing consists of remodelling grammatical "ways of saying things". Despite this fundamental difference, lexical and grammatical calquing share a single cause: bilingual speakers' need to express the same meaning in two languages (**Sasse1992**). This also means that everything that expresses meaning (i.e. morphemes, lexemes, and constructions) can, in principle, be a source of calquing.

Focusing mainly on the transfer of linguistic matter, Van Coetsem (**Coetsem1988**) does not overtly mention the possibility of transferring lexical and grammatical meanings through calquing. This chapter thus aims at relating contact induced changes produced by calquing to the principle of language dominance as postulated by Van Coetsem.

## 2 Contact-induced changes and calquing

### 2.1 Lexical calquing

According to Haspelmath (2009, 39), a lexical calque is a lexical unit that was created by an item-by-item translation of the source unit. This type of contact-induced change occurs as bilingual speakers reorganise the lexicon of one of their languages to match the semantic organisation of the other (**Ross2007**). Adopting the psycholinguistic standpoint of language dominance, Winford (2003, 345) regards lexical calquing as a subtype of lexical borrowing, which is a combination of recipient language (RL) lexemes in imitation of source language (SL) semantic patterns. In contrast, I will show that, though lexical calquing can easily be triggered by RL-dominant speakers, it can also be a product of imposition via SL agentivity. In order to do this, I will mainly focus on calquing of compound nouns. A compound noun is here defined as a series of two or more lexemes, which is semantically conceived as a single unit. Each component of the compound can function as a lexeme independent from the other(s), and may show some phonological and/or morphological constraints within the compound when compared to its isolated syntactic usage (**Bauer2001**). Against this backdrop, I will specifically discuss noun–noun compounds as they represent the more uniform phenomenon of nominal compounding in the world's languages (Pepper Forthcoming) As we will see, the transfer of the semantics of compound nouns does not imply any morphosyntactic change in Arabic, as calqued compounds are typically adjusted to fit RL morphosyntactic patterns.

Generally speaking, lexical calquing through borrowing can occur in indirect contact situations characterized by a very low degree of bilingualism. This is be-

cause RL monolinguals can also be agents of lexical borrowing (**Coetsem1988**) Typical instances of lexical calquing via RL agentivity are related to the transfer of the semantic patterns of English compound nouns in modern Arabic dialects. This kind of transfer is linked to the expansion of the non-core Arabic lexicon for expressing previously unknown concepts. A prime example is the English calque *lōḥit il-mafatīḥ* 'keyboard' (Lit. 'the board of keys') in Egyptian Arabic (**WilmsenWoidich2009**). Here, it can be clearly seen that the transfer of the semantic organization of the SL compound noun does not affect the morphosyntax of the RL, as the word order of the English nominal juxtaposition is reversed to fit the Arabic construct state.

Lexical calquing can also take place in prolonged contact situations, as testified by numerous Italian compounds in Maltese. A singular case of mixed calquing is that of *wiċċ tost* 'shameless person' (Lit. 'tough face') deriving from the Italian compound *faccia tosta* 'shameless person' (Lit. 'tough face') *(Falzon2013)*. On the one hand, the first lexical item of the compound presents an Arabic phonological form while expressing semantic properties associated with the lexeme 'face' in Italian. On the other hand, the second lexical item clearly results from the borrowing of the adjective *tosto* 'hard, tough' retaining both the Italian phonological matter and semantic properties. The mixed nature of this compound brings to the fore the complementary relationship between RL and SL agentivity and shows that it is not always a trivial matter to distinguish between imposition and borrowing. However, Maltese also gives evidence of genitive compounds in which both lexical components have an Arabic phonological form coupled with Italian semantic properties. This is the case of the compound nouns *saba' ta' sieq* 'toe' calqued on the Italian *dito del piede* 'toe' (Pepper Forthcoming). Such instances of lexical calquing clearly mirror semantic properties of SL lexemes and they most plausibly result from borrowing via RL agentivity (cf. Ćeplö & Lucas, this volume: §2.2.1).

Ḥassāniyya Arabic, for its part, presents many compound nouns that are traditionally analysed in terms of substratum interference from Zenaga Berber (Taine-**Cheikh2008**, 2012). Also in this case, the transfer of the semantic properties of the SL does not produce any morphosyntactic change in Arabic, as we can see in the following pairs of examples:

(1)  Ḥassāniyya Arabic (Taine-**Cheikh2008**: 126)
      kṛaʕ lə-ɣṛab
      foot DEF-crow
      'aquatic herbaceous plant' (Lit. 'crow's foot')

(2) Zenaga Berber (Taine-**Cheikh2008**: 126)
aðaʔṛ ən  tayyaḷ
foot GEN crow
'aquatic herbaceous plant' (Lit. 'crow's foot')

(3) Ḥassāniyya Arabic (Taine-**Cheikh2008**, 126)
sayllāl lə-ʕrāgib
ripper DEF-ankle.PL
'honey badger' (Lit. 'ripper of ankles')

(4) Zenaga Berber (Taine-**Cheikh2008**, 126)
aməssäf ən  ūržan
ripper GEN ankle.PL
'honey badger' (Lit. 'ripper of ankles')

Taine-Cheikh (2008, 115) stresses that it is somewhat difficult to trace back the origin of these compounds. Accordingly, she speaks of a process of convergence between the two languages, rather than determining the direction of the semantic transfer. However, it should be observed that these compound nouns are not attested in other spoken varieties of Arabic. Furthermore, since at least the mid-twentieth century, Berbers in Mauritania have been gradually loosing competence in Zenaga, in favour of Arabic (Taine-**Cheikh2012**, 100), while Zenaga is rarely acquired as second language by Ḥassāniyya Arabic speakers. In such a context, the most probable agents of contact-induced change were former Berber-dominant speakers who gradually shifted to Arabic. Thus, it seems plausible that the transfer of the semantic properties of Zenaga compounds has been achieved through imposition, rather than through borrowing.

Nigerian Arabic also shows interesting instances of lexical calquing as a consequence of a longstanding contact with Kanuri, a Nilo-Saharan language widely spoken in the Lake Chad area. Owens (2015, 2016) gives evidence of the transfer of the semantic properties of numerous compound nouns including the lexeme *ṛās* 'head'. Similar to the previous instances of compound calquing, the integration of Kanuri semantic patterns does not affect the Arabic morphosyntax, as we can see in the following pairs of examples:

(5) Nigerian Arabic (**Owens2016**, 69)
ṛās  al-bēt
head DEF-house
'roof' (Lit. 'head of house')

(6)  Kanuri (**Owens2016**, 69)
    kəla  fato-be
    head house-GEN
    'roof' (Lit. 'head of house')

(7)  Nigerian Arabic (**Owens2016**, 65)
    ṛās   al-qalla
    head  DEF-corn
    'tassel' (Lit. 'head of corn')

(8)  Kanuri (**Owens2016**, 65)
    kəla  argəm-be
    head corn-GEN
    'tassel' (Lit. 'head of corn')

According to Owens (2016, 65), Kanuri–Arabic bilingualism, with Arabic being a minority language, would have been the foremost factor underlying the transfer of these compound nouns into Nigerian Arabic. He further stresses that Kanuri is the main source of compound nouns in a number of other minority languages in the area (e.g. Kotoko, Glayda, and Fulfulde) and that there is little evidence of Kanuri to Arabic shift in the region (**Owens2014**, 147). However, the fact that Kanuri represents the majority language of northeastern Nigeria, does not shed light on the transfer mechanism lying behind lexical calquing in Nigerian Arabic. This is because speakers can be linguistically dominant in a socially subordinate language (**Winford2005**, 376). In fact, such contact settings are closely tied to SL agentivity, as the youngest bilingual generations tend to impose semantic features from their dominant language (i.e. Kanuri) onto the ancestral language (i.e. Arabic). It is only at a later stage that these innovations are borrowed by older bilingual speakers who are still dominant in Arabic.

The fact that Nigerian Arabic speakers have gradually developed a high bilingual proficiency in Kanuri is also testified by the transfer of a number of idiomatic expressions. In this regard, Ross (2007, 122) observes that calquing of meaning is not only reflected in word compounding, but also in lexical collocations of idiomatic expressions. These are combinations of lexical items that are semantically idiosyncratic as they have a pairing of form and meaning that cannot be predicted from the rest of the grammar. Examples (9)–(10) provide evidence of an idiomatic Kanuri calque in Nigerian Arabic.

(9)  Nigerian Arabic (Ritt-**BenmimounEtAl2017**, 77)

šuqul      šāl           ṛās-i
something carry.PRF.3SG.M head-OBL.1SG

'Something distracted me.' (Lit. 'Something carried my head.')

(10)   Kanuri (Ritt-**BenmimounEtAl2017**, 77)
       awo-de     kəla gō-zə-na
       something head carry-3SG-PRF

       'Something distracted me.' (Lit. 'Something carried head.')

   Given that idiomatic expressions are syntactically compositional (i.e. their lexical components behave syntactically as they do in non-idiomatic expressions), it is not only the meanings expressed by the lexeme 'head' which correspond between Nigerian Arabic and Kanuri, but also their idiomatic collocations, which align between the two languages (**Owens2014**, 157). Besides, it is worthwhile noting that also idiomatic expressions are adjusted to fit RL morphosyntactic patterns. This is evidenced by the inalienable possession of body parts in Nigerian Arabic (*ṛās-i* 'my head'), which is instead unattested in the SL (*kəla* 'head'). Even if we cannot exclude the possibility that these kinds of calques are a product of borrowing, it is evident that their integration needs a high proficiency in the SL for individuating the single idiomatic collocations of lexical items. Furthermore, differently from borrowed calques, imposed idiomatic expressions can significantly affect the lexical semantics of the RL created by SL-dominant bilinguals and thus produce grammatical changes in the long run.

   Finally, lexical calquing via SL-agentivity can also take place in extreme contact situations such as creolization. For instance, Juba Arabic, the Arabic-based pidgincreole spoken in South Sudan, shows numerous calques in which Arabic-derived lexemes are compounded according to the semantic patterns of Bari, the main substrate language of Juba Arabic (**Manfredi2017**, 50; **Nakao2012**). As we can see in (11)–(12) and (13)–(14), the word order in Juba Arabic compounds follows the order of Bari compounds. However, this cannot be seen as an innovative morphosyntactic development, as the possessed–possessor order matches also with the Arabic lexifier.

(11)   Juba Arabic (**Nakao2012**, 136)
       éna ta    séjera
       eye GEN tree

       'fruit' (Lit. 'eye of tree')

(12)   Bari (**Nakao2012**, 136)

koŋe lo-ködini
eye GEN-tree
'fruit' (Lit. 'eye of tree')

(13) Juba Arabic (**Nakao2012**, 137)
ída    ta    fil
hand GEN elephant
'trunk' (Lit. 'hand of elephant')

(14) Bari (**Nakao2012**, 137)
könin lo-tome
hand  GEN-elephant
'trunk' (Lit. 'hand of elephant')

Given that the asymmetric contact situation leading to creole formation limits access to the superstrate language (i.e. Sudanese Arabic), the semantic patterns of substrate languages (i.e. Bari) can be easily carried over into the creole in ways peculiar to imposition via SL-agentivity.

All things considered, unlike lexical borrowing, lexical calquing allows for a semantic overlapping of RL and SL lexical entries and it can also produce important structural changes.

## 2.2 Grammatical calquing

Grammatical calquing brings about a match between the grammatical categories of two languages and the memberships of these categories (**Ross2007**, 132). **HeineKuteva2005** suggest that the grammatical changes induced by calquing can be better analysed in terms of contact-induced grammaticalization (see also Leddy-Cecere, this volume). In fact, the calquing of the semantic properties of lexical and grammatical items may lead to the grammaticalization of innovative syntactic structures in the RL matching with those of the SL. From the traditional sociohistorical perspective of contact-induced change (**ThomasonKaufman1988**), grammatical calquing is basically seen as a product of language shift. In contrast, Ross (2007, 131) argues that grammatical calques can widely occur in situations of language maintenance. Actually, the different grammatical outputs of calquing mainly depend on the way in which they are transferred from the SL into the RL and, by extension, on different kinds and degrees of bilingualism.

For the sake of this chapter, I distinguish between three different types of grammatical calquing:

- Calquing of polyfunctionality of lexical items without syntactic change;

- Calquing of polyfunctionality of grammatical items inducing syntactic change;

- Narrow syntactic calquing (without calquing of polyfunctionality of lexical/grammatical items).

Being lexical in nature, the first of these three types of grammatical calquing can be triggered by both imposition via SL agentivity and borrowing via RL agentivity, whereas the two latter types are likely to result only from imposition via SL agentivity.

Calquing of polyfunctionality patterns of lexical items is by far the most common type of grammatical calquing, and it can be exemplified by the comparison of reflexive anaphors in different Arabic dialects. As is well known, Classical and Standard Arabic express a reflexive meaning either by means of agent-oriented derived verbs lacking an overtly expressed patient (e.g. *istaḥamma* 'he washed himself') or by anaphoric constructions in which the syntagm *nafs*-PRO.OBL 'soul-PRO.OBL' marks coreferentiality between the agent and the patient of the predicate (e.g. *qatala nafsa-hu* 'he killed himself'). Nevertheless, as a result of contact with different languages, a number of modern Arabic dialects have grammaticalized other lexical sources for expressing a reflexive meaning. Western Maghrebi dialects are a case in point. As we can see in (15)–(16), both Moroccan and Ḥassāniyya Arabic have grammaticalized the nominal syntagm *ṛāṣ*=PRO.OBL 'head-PRO.OBL' as default reflexive anaphor.

(15) Moroccan Arabic (D. Caubet, personal communication)
qtəl       ṛās-o
kill.PRF.3SG.M head-3SG.M
'He killed himself.' (Lit. 'He killed his head.')

(16) Ḥassāniyya Arabic (Taine-**Cheikh2008**, 16)
ktəl       ṛāṣ=u
kill.PRF.3SG.M head-3SG.M
'He killed himself.' (Lit. 'He killed his head.')

This reflexive use of the lexeme 'head' has generally been interpreted as substrate interference from Berber languages (El **Aissati2007**, 197), in which the same grammaticalization path is attested, as shown in the following examples from Tarifit and Zenaga:

(17) Tarifit Berber (**Kossmann2000**)
   γətšaθ        ixəf nnəs
   beat.PRF.3SG.M head POSS.3SG.M
   'He beats himself.' (Lit. 'He beats his head.')

(18) Zenaga Berber (Taine-**Cheikh2008**)
   yəʔna         iʔf-ən-š
   kill.PRF.3SG.M head-GEN-3SG.M
   'He killed himself.' (Lit. 'He killed his head.')

The lexeme for 'head' is the second most common source of grammaticalization of reflexive anaphors worldwide (König, Siemund, and **Töpper2013**) and its occurrence is particularly common in West Africa (**Heine2011**, ). In this scenario, it should be stressed that the reflexive function of the lexeme 'head' is an innovative feature of both Arabic and Berber varieties of northwestern Africa. By way of illustration, other Berber languages typically use the reflexive anaphor *iman*-POSS 'soul-POSS', as we can see in the following example from Kabyle.

(19) Kabyle (**Mettouchi2012**)
   n-səlk-dd          iman-ntə
   1PL-spare.PRF.PROX soul.ABS.SG.M-POSS.1PL.F
   'We saved ourselves.'

In addition, the known Arabic–Berber contact situation, in which second language learners of Berber only played a marginal role in triggering contact-induced change in Arabic, suggests that the contact induced grammaticalization of 'head' in westernmost Arabic dialects resulted from an imposition enacted by former Berber-dominant speakers.

A similar instance of calquing in the domain of anaphoric reflexive constructions is found in Kordofanian Baggara Arabic, a western Sudanic dialect spoken in the Nuba Mountains area, in central Sudan. In this case, the source of the reflexive anaphor is the lexeme for 'neck', as we can see in (20).

(20) Kordofanian Baggara Arabic (**Manfredi2010**, 176)
   abrahīm gaṣṣa      ragabt-a
   Ibrahim  cut.PRF.3SG.M neck-3SG.M
   'Ibrahim cut himself.' (Lit. 'Ibrahim cut his neck.')

Different from 'head', the grammaticalization of 'neck' as a reflexive anaphor is quite rare in Africa (**Heine2011**), but it is attested in a number of Niger-Kordofanian

languages spoken in the same region. Such is the case of Tagoi (21) and Koalib (22).

(21) Tagoi (**Alamin2015**)
t-ágám   t-ùrúŋ   ínní
NC-neck NC-POSS.3 kill.PRF.3
'He killed himself.' (Lit. 'He killed his neck.')

(22) Koalib (N. Quint, personal communication)
ɛɽŋɛ r-ɔkwɽɔ  r-ùŋwún
kill   NC-neck NC-POSS.3
'To kill oneself.' (Lit. 'To kill his neck.')

Similar to the situation described with reference to western Maghrebi dialects, Arabic-speaking groups in the Nuba Mountains have hardly developed any bilingual competence in local Niger-Kordofanian languages. Therefore, it seems likely that the calquing of the polyfunctionality patterns of 'neck' has been imposed by Arabized populations who were dominant in the SL.

Maltese also provides remarkable examples of calquing of polyfunctionality of lexical items. This is particularly evident in the domain of auxiliary verbs (**Vanhove1993**; Vanhove, Miller, and **Caubet2009**). A well-known example is that of the lexical verb *ġie* 'come' used as an auxiliary for expressing a dynamic passive (23) in the same way as Italian (24).

(23) Maltese (Borg and Azzopardi-**Alexander1997**, )
it-tabib   ġie        afdat   bi-l-każ
DEF-doctor come.PRF.3SG.M trusted with-DEF-case
'The doctor was entrusted with the case.' (Lit. 'The doctor came entrusted with the case.')

(24) Italian (own knowledge)
non  venne         creduto
NEG come.PRF.3SG.M trusted
'He was not trusted.' (Lit. 'He did not come trusted.')

Even if imposition played a role in the emergence of Maltese (Ćeplö & Lucas, this volume), it is generally accepted that intertwined languages emerge mainly from a widespread process of borrowing in Van Coetsem's terminology (**Winford2005**, ; **Manfredi2018**). This suggests that, different from the aforementioned grammaticalization of reflexive anaphors in Arabic dialects, the calquing

of polyfunctionality of lexical verb 'come' in Maltese was most likely triggered by agentivity of RL dominant speakers.

Regardless of the different contact situations, what holds all the previous instances of grammatical calquing together is the fact that the transfer of patterns of grammaticalization did not produce any syntactic change in Arabic. In contrast to the above, the calquing of polyfunctionality of grammatical items can be accompanied by important typological changes. This is the case of the grammaticalization of prototypical passive constructions in Juba Arabic (**Manfredi2017**, 92, 2018, 415). As we can see in (25), the South Sudanese pidgincreole presents an innovative passive construction in which the patient occupies the syntactic slot of a preverbal subject, whereas the oblique-marked agent is introduced by the comitative preposition *ma-* 'with'.

(25)  Juba Arabic (**Manfredi2017**)
      bab  de      kasurú   ma-jón
      door PROX.SG break.PASS with-John

      'This door has been broken by John.' (Lit. 'This door has been broken with John.')

Interestingly, this prototypical passive construction is not attested in the lexifier language of Juba Arabic (i.e. Sudanese Arabic), which instead makes use of impersonal passive constructions with a default 3PL.M subject.

(26)  Sudanese Arabic (S. Manfredi, own knowledge)
      kassaru-hu
      break.PRF.3PL.M-3SG.M

      'It got broken.' (Lit. 'They have broken it.')

Indeed, the grammaticalization of this complex syntactic structure is the result of the calquing of the functional properties associated with the comitative preposition of the main substrate language, Bari. Bari presents the same kind of prototypical passive construction in which an oblique-marked agent is introduced by the preposition *ko-* 'with'.

(27)  Bari (**Owen1909**)
      niena    wuret a-wur-ö            ko-nan
      PROX.SG book   3SG.PAST-write-PASS with-1SG

      'This book has been written by me.' (Lit. 'This book has been written with me.')

If we assume that the emergence of creole languages is always induced by the disruption of the transmission of the lexifier language (**Comrie2011**), we can conclude that Bari speakers have imposed the semantics of their dominant language on a grammatical item derived from Arabic, and thus induced profound changes in the word order of the creole when compared to its lexifier language.[1] In light of the above, the contact dynamics lying behind the calquing of polyfunctionality of grammatical items are quite restrictive as they are most likely a product of imposition via SL agentivity.

The third kind of grammatical calquing is linked to the transfer of syntactic patterns without transfer of polyfunctionality of either lexical or grammatical items. This narrow type of syntactic calquing can be exemplified by possessor doubling in Central Asian Arabic (**Ratcliffe2005**). Clitic doubling is a construction in which a clitic co-occurs with a full nominal phrase in argument position, forming a discontinuous constituent with it. Various forms of clitic doubling have arisen in a number of Arabic varieties as a result of contact with different substrate/adstrate languages (**Souag2017**). In regard to possessive constructions, Arabic typically presents a possessed–possessor order. In contrast, Central Asian Arabic (28) gives evidence of the opposite order with obligatory possessor doubling in the same way as Tajik (29).

(28)  Central Asian Arabic (**Ratcliffe2005**; **Souag2017**)
      amīr    wald-u
      prince son-3sg.m
      'the prince's son'

(29)  Tajik (**Souag2017**)
      buxoro   universitet-ash
      Bukhara university-3sg
      'Bukhara University'

Souag (2017, 157) states that double possessor constructions in Central Asian Arabic are instances of grammatical calquing, accommodated through the reinterpretation of pre-existing topicalized constructions. This means that, different from the syntactic changes induced by the calquing of polyfunctionality of morphemes, the emergence of double possessor constructions in Bukhara Arabic would have been favoured by a formal congruence between SL and RL syntactic

---

[1]This kind of syntactic change accompanied by the calquing of semantic properties of substrate items in creole languages is traditionally labelled 'relexification' (**Lefebvre1998**).

structures. As such, this instance of contact-induced morphosyntactic restructuring (i.e. metatypy) does not derive from a direct copying of a double possessor construction. Rather, it consists in speakers expressing a possessive meaning in Arabic by using a construction which they equate with the construction in adstratal languages (**Ross2007**). If we consider that the youngest speakers of Central Asian Arabic are gradually losing competence in their ancestral language in favour of socially dominant languages (**Chikovani2005**), it is plausible to think that such kind of syntactic restructuring can only be a result of imposition via SL-agentivity. Still, given our limited diachronic knowledge, we cannot exclude the hypothesis of an early process of borrowing enacted by former Arabic-dominant speakers.

# 3 Conclusion

Van Coetsem (1988, 20) suggests that the variable outcomes of language contact are primarily a reflex of the "stability gradient" of language, which induces speakers to preserve the domains of their dominant language that are less affected by change. As lexicon is the most unstable linguistic domain, it is likely to be transferred via RL-agentivity. In contrast, morphosyntax and phonology are considered to be relatively stable domains and they are expected to be transferred only via SL-agentivity. Against this background, it is unclear how the transfer of sematic features deprived of morphophonological matter should be understood in relation to the linguistic dominance of the agents of contact-induced change.

If we look at the previously analysed instances of lexical calquing (§2.1), it is evident that the transfer of the semantic features of nominal compounds can take place within speech communities with a very low degree of bilingualism, as in the case of Egyptian-Arabic-dominant speakers borrowing the semantics of English compounds. But it is also true that compound calquing can be a product of imposition resulting from ongoing language shift or pidginization, and the transfer of semantic features of single lexical items within idiomatic expressions always requires a widespread proficiency in the SL, as in the case of Arabic–Kanuri bilingualism in northern Nigeria.

As far as grammatical calquing is concerned (§2.2), I have shown that calquing of the polyfunctionality of lexical items can be triggered either by imposition, as in the case of substrate interference in Ḥassāniyya and Baggara dialects, or by borrowing in the emergence intertwined languages such as Maltese. Calquing of polyfunctionality of grammatical items, for its part, requires a higher degree of linguistic abstraction for the identification of a functional overlap between

morphemes. Accordingly, this type of transfer will typically occur via imposition by SL-dominant speakers in deep contact situations such as creolization. In the same manner, narrow syntactic calquing requires high bilingual proficiency, as it necessitates the recognition of some formal congruence between the SL and the RL, as shown by the emergence of possessor doubling in Central Asian Arabic.

To stay somewhat in line with the stability gradient principle, we could argue that, in absence of the transfer of linguistic matter, the semantic properties of morphemes and syntactic constructions are more stable than those of lexical items. However, such a generalization would be misleading without an in-depth knowledge of the sociolinguistic circumstances underlying a specific instance of second language acquisition (i.e. symmetric bilingualism, asymmetric bilingualism, multilingualism, pidginization/creolization). Thus, it becomes evident that the recognition of different patterns of bilingualism within the same community remains the only way to identify the transfer type at play in a given contact situation, regardless of its different structural outputs.

Drawing on the available literature, this chapter has surveyed only a few instances of lexical and grammatical calquing induced by contact between Arabic and other languages. This is mainly because we lack information about calquing in dialect contact situations. Indeed, it is regrettable that studies dealing with dialect contact and new dialect formation are still exclusively focused on the diffusion of few lexical and morphophonological features, while disregarding the transfer of semantic and syntactic patterns. Fine-grained analyses of calquing in dialect contact situations thus remain a major desideratum for the development an aggregate variationist Arabic dialectology.

## Further reading

**Keesing1988** adopts the notion of calquing and describes the transfer of semantic properties of Oceanic morphemes in Melanesian Pidgin.
**Meyerhoff2009**, by focusing on the notions of replication, transfer, and calquing, strengthens connections between variationist sociolinguistics and contact linguistics.
**Zuckermann2009** provides numerous instances of calquing in Modern Hebrew and analyses them in the light of the Congruence Principle.

# Abbreviations

ABS  absolute state                 NOM   nominative
DEF  definite article               NOM   nominative
1, 2, 3 first, second, third person
ABS absolute state
DEF definite article
F feminine
GEN genitive case and exponent
M masculine
NC noun class
OBL oblique
PASS passive
PAST past
PL plural
POSS possessive pronoun
PRG pragmatic marker
PRF perfect
PRO pronoun
PROX proximal
REFL reflexive
SG singular

# Chapter 10

# Arabic pidgins and creoles

??

??

> The chapter is an overview of contact-induced changes in eight Arabic-lexifier pidgins and creoles: Turku, Bongor Arabic, Juba Arabic, Kinubi, Pidgin Madam, Jordanian Pidgin Arabic, Romanian Pidgin Arabic, and Gulf Pidgin Arabic. The examples illustrate a number of selected features of these varieties. The focus is on two types of transfer, imposition and borrowing, within the framework outlined by van Coetsem (1988; 2000; 2003) and Winford (2005; 2008).

## 1 Introduction

This chapter aims to illustrate the emergence of Arabic-lexifier pidgins and creoles for which the contact situation – i.e. socio-historical context, the agents of change, and the languages involved – is at least relatively well known.

The varieties considered can be classified in two groups, in geographical, historical and developmental terms: the Sudanic pidgins and creoles, and the immigrant pidgins in various Arab countries. Geographically, the Sudanic varieties developed in Africa – in present-day South Sudan, Chad, Uganda, and Kenya. Historically, the Sudanic varieties derive from a putative common ancestor, a pidgin that emerged in southern Sudan, in the first half of the 19<sup>th</sup> century. Various Turkish-Egyptian military expeditions between 1820 and 1840 opened southern Sudan for the slave trade. Permanent camps were set up soon after by slave traders in the White Nile Basin, Bahr el-Ghazal and Equatoria Province, inhabited by an Arabic-speaking minority and a huge majority of slaves from various ethnic and linguistic backgrounds. **After1850**, the slave traders' settlements were turned into military camps in which a military pidgin emerged, which is traditionally referred to as "Common Sudanic Pidgin Creole Arabic" (**ToscoManfredi2013**:

253). Two subgroups of Sudanic varieties are recognized: the western branch, consisting of Turku and Bongor Arabic (in Chad), and the eastern one, made up of Juba Arabic (in Sudan) and Kinubi (spoken in Uganda and Kenya).

Immigrant pidgins emerged in the eastern part of the Arab World, in Lebanon, Jordan, Iraq and the countries of the Arab Gulf. Historically, these do not go back to more than 50 years. All these varieties are incipient pidgins.

The contact situations illustrated presuppose: (i) a SL and a RL; (ii) agents of contact-induced change, who may be either SL or RL speakers; (iii) a psycholinguistically dominant language, which is not necessarily a socially dominant language (van **Coetsem1988**; 1995; 2000; 2003; **Winford2005**; 2008). A distinction is made between two types of transfer: imposition and borrowing (van **Coetsem1988**; 2000; 2003). Imposition involves SL-dominant speakers as agents (SL agentivity), is typical of second language acquisition, and induces changes mostly in phonology and syntax, although it may also include transfer of lexical items from the dominant SL into the non-dominant RL (van **Coetsem1995**: 18; **Winford2005**: 376). Borrowing normally involves RL-dominant speakers as agents (RL agentivity), it typically targets lexical items, but may also include transfer of morphological material from a non-dominant SL into the dominant RL.

In light of their sociolinguistic history, the varieties considered all emerged under conditions of untutored, short term, second language acquisition by adults dominant in their socially subordinate SLs. Second language acquisition, *a fortiori* with adults, triggers processes such as imposition via SL agentivity (i.e. substrate influence), simplification (**Trudgill2011**: 40, 101) – also known as restructuring (**Lucas2015**), as well as language-internal – i.e. non-contact-induced – developments such as grammatical reanalysis (**Winford2005**).

As in **Manfredi2018**, the focus of this chapter is on imposition and borrowing. It does not illustrate restructuring which does not involve any kind of transfer, but often involves a reduction in complexity (**Lucas2015**). In the case of Arabic pidgins and creoles, restructuring is manifest in the domain of morphology, in e.g. the loss of the Arabic verbal affixes and of the nominal and verbal derivation strategies (**Miller1993**).

The examples are illustrative only of selected contact-induced features of Arabic pidgins and creoles and their number has been kept at a reasonable minimum. The examples from Arabic and the pidgins and creoles considered appear in a uniform system of transliteration.

The chapter is organized as follows. §§**??** to 5. are concerned with the Sudanic pidgins and creoles. §§**??** to 9. focus on immigrant varieties. §**??** summarizes the

findings and briefly discusses issue for further research.

## 2 Turku and Bongor Arabic

### 2.1 Current state and historical development

Turku is an extinct pidgin, formerly spoken in the Chari-Bagirmi region in western Chad (**Muraz1926**). After the abolition of slavery by the Turkish-Egyptian government in 1879, the Nile Nubian trader Rabeh withdrew with his slave soldiers into Chad. From a socio-linguistic point of view, Turku was initially a military pidgin. However, it later became one of three trade languages in what was then French Equatorial Africa, along with Sango and Bangala (**ToscoOwens1993**: 183). Turku was a stable pidgin, which does not appear to have creolized (**ToscoOwens1993**).

Bongor Arabic is spoken in southwestern Chad, in and around the town of Bongor, the capital of the Mayo-Kebbi Est region, close to the border with Cameroon (**Luffin2013**). Given the many structural features it shares with Turku, it is plausible to assume that Bongor Arabic developed from the former. Sociolinguistically, Bongor Arabic is a trade pidgin, used by the local Masa and Tupuri populations with Arabic-speaking traders. It is currently a stable pidgin, but it exhibits features indicative of depidginization under the influence of Chadian Arabic. No information about the number of speakers is available.

### 2.2 Contact languages

The lexifier language of Turku and Bongor Arabic is Western Sudanic Arabic. The substratal input was provided by languages of various genetic affiliations: Nilo-Saharan – e.g. Bagirmi, Mbay, Ngambay, Sar, Sara (Central Sudanic), Kanuri (Western Saharan); Afro-Asiatic – Hausa (West Chadic); Niger-Congo – Fulfulde. In the case of Turku an additional contributor was the creole language Sango. Both in Turku and in Bongor Arabic there is also adstratal input from French. The adstrate of Bongor Arabic additionally includes two languages: Masa (Nilo-Saharan, Western Chadic) and Tupuri (Niger-Congo).

### 2.3 Contact-induced changes

#### 2.3.1 Phonology

The substrate languages do not have /x/, which is generally replaced by /k/: Turku *kamsa* 'five' < CA *ḥamsa*; Bongor Arabic *kídma* 'work' < CA *ḫidma*. Many of the substrate languages do not have /f/, which is substituted with /p/ or perhaps

/ɸ/[1], e.g. Turku **pf**il 'elephant' < CA **f**īl. In French loanwords, the reflexes of /v/ are either /b/ or /w/: Bongor Arabic **bot**é 'vote (v)' < French **v**oter, **w**otír 'car' < French **v**oiture.

The consonants ɲ/ and /ŋ/ occur only in loanwords: Turku konpa**ny**e 'company' < French compag**n**ie, **ng**ari 'manioc' < Mbay **ng**àrì; Bongor Arabic **ng**ambáy 'Ngambay' < Ngambay **ng**àmbáy; /v/ and /ʒ/ occur only in phonologically non-integrated words of French origin, Turku si**v**il 'civilian' < French ci**v**il; Bongor Arabic **ž**urnalíst < French **j**ournaliste.

Variation affects several consonants. For instance, [f] occurs in variation with [b] or [p]: Turku **f**išan ~ **b**išan 'because'; Bongor Arabic má**f**i ~ má**p**i 'neg' < CA mā **f**í, so**f**ér ~ so**p**ér 'driver' < French chau**ff**eur. Most of the substrate languages do not have /ʃ/, which accounts for [ʃ] ~ [s] variation, in words with either etymological /s/ or /ʃ/: Turku gasi ~ gaʃi 'expensive' < CA gāsī, biriš ~ biris 'mat' < CA birīš; Bongor Arabic máʃi ~ mási 'go'. The usual reflexes of French /ʒ/, absent from the phonological inventories of the substrate languages, are /z/, /dʒ/ and /s/ respectively: Turku **ǧ**inenal 'general' < French **g**énéral, **s**uska 'until < French **j**usqu'à; Bongor Arabic **z**úska < French **j**usqu'à.

Finally, vowel length is not distinctive: Turku, Bongor Arabic kalam 'speech; speak' < CA kalām 'speech'.

### 2.3.2 Morphology

On current evidence (**Luffin2013**: 180–181), Bongor Arabic exhibits signs of depidginization under the influence of Chadian Arabic. The most striking instance of this is the use of pronominal suffixes, unique among Arabic-lexifier pidgins and creoles:

(1)  Bongor Arabic (**Luffin2013**)
     índi gáy árifu úsum-**i**
     2sg ipfv know name poss.1sg
     "You know my name."

Also, verbal affixes are sporadically used:

(2)  Bongor Arabic (**Luffin2013**)
     a.  ána ma **n**-árfa
     1sg neg 1sg know
     "I don't know."

---

[1]Given the transcription with <pf> (**Muraz1926**).

(3) b. anína rikíb-**na** wotír da  sáwa
     1PL ride  PFV      .1PL  car DEM together
     "We took the car together.'

These cases might be analyzed as borrowing under *sui generis* RL agentivity, whereby morphological material from a non-dominant SL is imported into a non-dominant (second) RL.

### 2.3.3 Lexicon

A part of the non-Arabic vocabulary of Turku can be traced back to its substrate languages (Avram in press). Most of the loanwords are from Sara-Bagirmi languages: *adinbang* 'eunuch' < Bagirmi *ádim mbàŋ* 'servant of the sultan'; *gao* 'hunter' < Sar *gáw*; *ngari* 'manioc' < Mbay *ngàrì*. The second most significant important contributor is Sango: *kay* 'paddle' < Sango *kâî*, *tipoy* 'carrying hammock' < *típóí*. A few words can be traced to Fulfulde and Kanuri: *kelkelbanǧi* 'golden beads' < Fulfulde *kelkel-banja*; *wélik* 'lightning' < Kanuri *wulak* 'flash of lightning'. In a number of cases, the exact SL cannot be established: *koporo* '0.10 Francs' < Fulfulde, Sango, Sara *koporo* 'coin'; *gurumba* 'hat' < Hausa *gurúmba*, Kanuri *gurumbá*. As for Bongor Arabic, its African adstrate languages have contributed only few loanwords, e.g. *bursdíya* 'Monday'. There are also loanwords from French. In Turku most of these relate to the military (**ToscoOwens1993**: 262–263), e.g. Turku *itenan* 'lieutenant' < French *lieutenant*, *permišon* 'permission' < French *permission*. In addition to nouns French loanwords include some verbs, e.g. Bongor Arabic *komandé* 'order' < French *commander*, and at least one functional word, Turku *suska*, Bongor Arabic *zúska* 'when, during' < French *jusqu'à* 'until'.

The substratal influence on Turku can also be seen in a number of compound calques (Avram in press; Manfredi this volume). Some of these are modelled on Sara-Bagirmi languages: *bahr gum* 'rising water', cf. Ngambay *màn-kàw*, lit. 'river goes'; *nugra ana asal* 'beehive', cf. Ngambay *bòlè-tənji*, lit. 'hole (of) honey'. Other calques have equivalents in several SLs, e.g. *nugra haǧer* 'cave', lit. 'hole mountain/stone', cf. Kanuri *kûl kau-be* lit. 'cavity mountain-of', Ngambay *bòlò-mbàl* lit. 'hole mountain', Sango *dûtênë* lit. 'hole stone'.

# 3 Juba Arabic and Kinubi

## 3.1 Current state and historical development

Juba Arabic is mainly spoken in South Sudan; there are also diaspora communities, mostly in Sudan and Egypt. Two main reasons make it difficult to estimate its number of speakers. Firstly, while Juba Arabic is spoken as a primary language by 47% of the population of Juba, the capital city of South Sudan, it is also used as a second or third language by the majority of the population of the country (**Manfredi2017**). Secondly, the long coexistence of Juba Arabic with Sudanese Arabic, its main lexifier language, has led to the emergence of a continuum ranging from basilectal, through mesolectal, to acrolectal varieties; delimiting acrolectal Juba Arabic from Arabic is no easy task, particularly in the case of the large diaspora communities in Khartoum and Cairo.

Juba Arabic emerged as a military pidgin. Sociolinguistically, it is today an inclusive identity marker for the ethnically and linguistically diverse population of South Sudan (**ToscoManfredi2013**: 507). Developmentally, Juba Arabic is a pidgincreole[2].

The Mahdist revolt, which started in 1881, eventually brought to an end the Turkish-Egyptian control over Equatoria, in southern Sudan. Following an invasion by Mahdist rebels, the governor fled to Uganda, accompanied by slave soldiers loyal to the central government. These soldiers subsequently became the backbone of the British King's African Rifles. While some of the troops remained in Uganda, others were moved to Kenya and Tanzania. This led to the dialectal division between Ugandan and Kenyan Kinubi. Like Juba Arabic, therefore, Kinubi started out as a military pidgin, then underwent stabilization and expansion. Today, however, Kinubi is the only Arabic-lexifier fully creolized variety, i.e. a native language for its entire speech community.

Kinubi is spoken in Uganda and in Kenya. The number of speakers of Kinubi is a matter of debate. Ugandan Kinubi was spoken by some 15,000 people, according to the 1991 census, and Kenyan Kinubi by an estimated 10,000, in 2005. However, other estimates put the combined number of speakers at about 50,000. The largest communities of Kinubi speakers are in Bombo (Uganda), Nairobi (the Kibera neighbourhood) and Mombasa (Kenya).

---

[2]A pidgincreole is "a former pidgin that has become the main language of a speech community and/or a mother tongue for some of its speakers" (**Bakker2008**).

## 3.2  Contact languages

The main lexifier language of Juba Arabic is Sudanese Arabic, with some input from Egyptian Arabic and Western Sudanic dialects as well. The substrate is represented by a relatively large number of languages, belonging to super-phylums, Nilo-Saharan and Niger-Congo. The former includes Eastern Sudanic languages, such as Bari, Lotuho (Eastern Nilotic), Acholi, Belanda Bor, Dinka, Jur, Nuer, Päri, Shilluk (Western Nilotic), Didinga (Surmic), and Central Sudanic languages, such as Avokaya, Baka, Bongo, Ma'di, Moru; the Niger-Congo super-phylum is represented by e.g. Zande and Mundu. The main substrate language is considered to be Bari, including its dialects Kakwa, Kuku, Pojulu, and Mundari[3].

Given its sociolinguistic history, Kinubi shares much of its substrate with Juba Arabic. However, the substrate of Ugandan Kinubi additionally includes Eastern Sudanic languages, such as Alur, Luo (Western Nilotic), and Central Sudanic languages such as Mamvu, Lendu and Lugbara (**Owens1997**: 161; **Wellens2003**: 207), spoken in Uganda. Unlike Juba Arabic, Kinubi also exhibits the effects of the adstratal influence exerted by two Bantu languages, Luganda – particularly in Ugandan Kinubi – and Swahili – particularly in Kenyan Kinubi. One other language that should be mentioned is English, official both in Uganda and in Kenya.

## 3.3  Contact-induced changes

### 3.3.1  Phonology

A number of consonants found in Arabic, but absent from the phonological inventories of the substrate languages are either deleted or substituted. Consider the reflexes of pharyngeals: **h***áfla* 'feast' < SA *ḥafla*; *árabi* 'Arabic' < SA *ʕarabī*. The pharyngealized consonants are replaced by their plain counterparts: **t***owíl* 'long' < SA *ṭawīl*; **d***ul* 'shadow' < SA SA *ḍull*; *súlba* 'hip' < SA *ṣulba*; *zúlum* 'anger (v)' < SA *ẓulum*. The velar fricatives of Arabic are always replaced by velar stops: **k***ábara* 'piece of news' < SA *ḫabar*; *šó***k***ol* 'work' < SA *šoġol*, **g***árib* 'west' < SA *ġar(i)b*.

As in Juba Arabic, the pharyngeals of Arabic are either replaced or lost in Kinubi (**Owens1985**: 10; **Wellens2003**: 209–212). The earliest records of Ugandan Kinubi[4] are replete with illustrative examples (**Avram2017b**): **h***aǧa* 'thing' < SA *ḥāǧa*, *aram* 'thief' < SA *ḥarāmī*, *līb* < 'play (v)' < SA *liʕib*. The pharyngealized consonants are replaced by their plain counterparts, as in these examples from

---

[3]Sometimes considered to be separate languages (**Wellens2003**).
[4]The main ones are **Cook1905**, **Jenkins1909**, **Meldon1913**, **OwenKeane1915**.

early Ugandan Kinubi: *ṯowil* 'long' < SA *ṯawīl*; *ḏulu* 'shadow' < SA *ḏull*, *hisiba* 'measles' < SA *ḥiṣba*; *zulm* 'anger (v)' < SA *ẓulum*. Like Juba Arabic, Kinubi substitutes velar stops for the Arabic velar fricatives. Consider the following early Ugandan Kinubi forms: *kidima* 'work' < SA *ḫidma*; *šokolo* 'work' < SA *šoġol*, *balago* 'commandment' < SA *balāġ* 'message'. Substratal influence also accounts for consonant degemination, given that the substrate languages "lack these in all but a few morphonologically determined contexts" (**Owens1997**).

Substratal influence can also be seen in the occurrence of certain consonants. Consider first /ɓ/ and /ɗ/: Juba Arabic *d'éngele* 'liver' < Bari *denggele*; Juba Arabic *b'ónǧo* 'pumpkin' < Bongo *b'onǧo*. The other consonants which occur only in loanwords from the substrate and/or adstrate languages are /p/, /v/, /ʧ/, /ɲ/, and /ŋ/: Kinubi *lípa* 'pay (v)' < Swahili -*lipa*; Kinubi *camp* 'camp' < English *camp*; Kinubi *víta* 'war' < Swahili *vita*; Juba Arabic *cam* 'food' < Acholi, Belanda Bor, Jur *čama*, Juba Arabic *čainiz* < English *Chinese*, Kinubi *chai* 'tea' < Swahili *chai*; Juba Arabic *nyékem*, Kinubi *nyékem* 'chin' < Bari *nyékem*, Kinubi *nyánya* 'tomato' < Swahili *nyanya*; Juba Arabic *ŋun* 'divinity' < Bari *ngun*. The integration of these phonemes is thus a result of borrowing (under RL agentivity) rather than of imposition.

The following instances of consonant variation are more common in Juba Arabic (**Manfredi2017**: 25–27). The most frequent one is [ʃ] ~ [s]: *geš* ~ *ges* 'grass'. Further, [z] is in variation with [ʤ] before /o/ and /a/: *zówǧu* ~ *ǧówǧu* 'marry', *záman* ~ *ǧáman* 'time; when'. There is also [p] ~ [f] variation in word-initial position, including in loanwords: *poǧúlu* ~ *foǧúlu* 'Pojulu', *prótestan* ~ *frótestan* 'Protestant'. Finally, the phoneme /f/ may also be phonetically realized as [p]: *nédifu* ~ *nédipu* 'clean (v)'. Of these cases of variation, the latter has been specifically attributed to substratal influence from Bari (**Miller1989**; **Manfredi2017**). It might be argued, however, that all these instances of consonant variation reflect the influence of the substrate languages, regardless of their genetic affiliations. The following do not have /ʃ/: Acholi, Avokaya, Baka, Bari, Belanda Bor, Bongo, Dinka, Jur, Lotuho, Ma'di, Moru, Mundu, Nuer, Päri, Shilluk, Zande. Of these Acholi, Belanda Bor, Bongo, Dinka, Jur, Nuer, Päri and Shilluk do not have /s/ either. A number of substrate languages do not have /z/: Acholi, Bongo, Belanda Bor, Dinka, Jur, Lotuho, Nuer, Päri, and Shilluk. All of these, however, have /ʤ/. Finally, /f/ is not part of the phonological inventory of Acholi, Bongo, Dinka, Jur, Nuer, Päri, and Shilluk, which do, however, have /p/. Given the intricacies of the distribution of /ʃ/, /s/, /z/, /ʤ/, /f/, and /p/ across the substrate languages, the types of variation illustrated are not surprising.

As in Juba Arabic, [ʃ] is in variation with [s] in Kinubi (**Owens1985**: 237;

**Owens1997**: 161; **Wellens2003**: 38; **Luffin2005**: 62; **Avram2017b**): early Ugandan Kinubi *šabaka ~ sabaka* 'net'). Although it is etymological /ʃ/ which is typically subject to variation, occasionally this also applies to etymological /s/: early Ugandan Kinubi *sikin ~ šekin* 'knife' < SA *sikkīn* (**Avram2017b**) and modern Kenyan Kinubi *fluš ~ flus* 'money' < SA *fulūs* (**Luffin2005**). Note that [ʃ] ~ [s] variation also extends to loanwords from Swahili (**Wellens2003**: 80; **Luffin2005**: 63; **Avram2017b**): early Ugandan Kinubi *šamba ~ samba* 'field' < Swahili **sh**amba. As Juba Arabic, Kinubi exhibits [z] ~ [ʤ] variation (**Owens1985**: 235; **Owens1997**: 161; **Wellens2003**: 215; **Luffin2005**: 63; **Avram2017b**): early Ugandan Kinubi *ǧalan ~ zalan* 'angry'. However, unlike Juba Arabic, in Kinubi the [z] ~ [ʤ] variation also occurs before the two front vowels /i/ and /e/: *ze ~ ǧe* 'as': early Ugandan Kinubi *anǧil ~ enzil* 'descend'. According to **Owens1997**, this "is due perhaps to Bari substratal influence, since Bari has only *j*, not *z*". In fact, as in the case of Juba Arabic, the same is true of several other substrate languages. Finally, there are instances of [l] ~ [r] variation (**Wellens2003**: 214; **Luffin2005**: 65), affecting both etymological /l/ and etymological /r/ in Arabic-derived words, e.g. *tále ~ táre* 'go out', *gerí ~ ge**l**í* 'near', and in borrowings, e.g. Ugandan Kinubi *čá**l**o ~ čá**r**o* 'village' < Luganda *e-kyalo*; Kenyan Kinubi *tumbí**l**i ~ tumbí**r**i* 'monkey' < Swahili *tumbi**l**i*. This variation seems to reflect the influence of Luganda and Swahili. In the former, [l] and [r] are in complementary distribution, with [r] occurring after the front vowels /i/ and /e/, and [l] elsewhere (**Wellens2003**), while in the latter [l] and [r] are in free variation (**Luffin2014**).

As in the substrate languages, there is no distinction between short and long vowels: Juba Arabic *sud**á**ni* 'Sudanese' < SA *sudānī*, Kinubi *kab**í**r* 'big' < SA *kab**ī**r*.

### 3.3.2 Morphology

Apart from the Arabic-derived plural suffixes *-at* and *-in*, Juba Arabic uses the plural marker of Bari origin *-ǧín* (**Nakao2012**; **Nakao2012**: 131; **Manfredi2014b**: 58), which is attached only to loanwords from local languages:

(4) Juba Arabic (**Manfredi2014b**)
    a.   kɔrɔpɔǧín < Bari kɔrɔpɔ + ǧín
    leaf PL
    "leaves"

(5)   b. *bengǧín* < Dinka *beng* + *ǧín*
    chief PL

*??*

"chiefs"

(6) c.    b'angiriǧín < Zande b'angiri + ǧín
cheek PL
"cheeks"

Another phenomenon worth mentioning is the occurrence in the speech of young, urban speakers of hybrid forms, which consist of the Bari relativizer *lo*- and a noun either from Arabic or from one of the substrate/adstrate languages (**Nakao2012**). Note, however, that there is a functional overlapping between Bari *lo*- and Sudanese Arabic *abu*.

(7) Juba Arabic (**Manfredi2017**)
a.   lobeléde < Bari lo + SA beled
REL country
"peasant"

(8) b.   lopómbe < Bari lo + Swahili pombe
REL alcohol
"drunkard"

Given that a relatively large number of Bari-derived words contain *lo*-(**Miller1989**; **Manfredi2017**: 46), the examples under (**??**) confirm the fact that morphological innovations are typically introduced through lexical borrowings via RL agentivity, and subsequently become productive in the RL.

Note, finally, that most of the speakers who use the plural marker *-ǧín* and the relativizer *lo*- are dominant in Juba Arabic. These cases therefore confirm the fact that RL monolinguals can be agents of borrowing (van **Coetsem1988**: 10).

A small number of Kinubi nouns borrowed from Swahili exhibit the Bantu nominal classifiers:

(9) Kinubi (**Wellens2003**)
a. ***mu-zé wa-zé***
CLS1-old man CLS2-old man
"old man" "old men"

do not put examples next to each other

(10) b.   **mu**-zukú    **wa**-zukú
CLS1-grandchild CLS-grandchild
"grandchild" "grandchildren"

(11)  c.              **m**zúngu        **wa**zúngu
      CLS1-European CLS2-European
      "European" "Europeans"

### 3.3.3 Syntax

Bureng **Vincent1986** first noted the similarity between the prototypical passive construction in Juba Arabic and its Bari counterpart:

(12)  Juba Arabic (Bureng **Vincent1986**: 77)
      Bágara áyinu  **ma**  Wáni
      cow     see.PST with Wani
      "The cow was seen by Wani."

(13)  Bari (Bureng **Vincent1986**: 77)
      Kítзŋ a   mɛtà kɔ`  Wànì
      Cow   PST see  with Wani
      "The cow was seen by Wani."

As can be seen, in both Juba Arabic and Bari the agent is introduced by the comitative preposition 'with'. This is a case of lexico-syntactic imposition via identification of SL and RL lexemes (**Manfredi2018**): the Juba Arabic lexical entry *ma* is derived from Sudanese Arabic *maʕ*, but its semantics reflects the influence of Bari *kɔ`*. The same is true of Kenyan Kinubi:

(14)  Kinubi (**Luffin2005**)
      yal-á    al akulú  **ma**  nas tomsá
      child-PL REL eat.PASS with PL  crocodile
      "children who were eaten by a crocodile"

Consider next the syntax of numerals in Kinubi (**Wellens2003**: 90; **Luffin2014**: 309). Their post-nominal placement is modelled on Swahili:

(15)  Kinubi (**Luffin2014**)
      wéle **kámsa** ma   baná  **árba**
      boy  five     with girl.PL four
      "five boys and four girls"

(16)  (10) Swahili (**Luffin2014**)

miti **mia** **tatu**
tree hundred three
"three hundred trees"

With cardinal numerals, the order is hundred + unit and thousand + unit respectively:

(17) Kinubi (**Luffin2014**)
elf     wáy
thousand one
"one thousand"

Kinubi thus follows the Swahili model:

(18) Swahili (**Luffin2014**)
elfu    moja
thousand one
"one thousand"

Consider also a case of syntactic change induced by lexical calquing. Juba Arabic *(fu)wata* 'ground' functions as an impersonal subject in weather expressions:

(19) Juba Arabic (**Nakao2012**)
**(fu)watá** súkun
ground   hot
"It is hot."

**Nakao2012** shows that this is also the case in Acholi and Ma'di:

(20) Acholi (**Nakao2012**)
**piiny** lyeet
ground warm
"It is warm."

(21) Ma'di (**Nakao2012**)
**vu**    aci
ground hot
"It is hot."

In fact, this type of sentences is widespread in the Western Nilotic substrate languages, such as Dinka, Jur, Päri, and Shilluk:

(22)  Dinka (**Nebel1979**)
      **Piny**  a-    tuc
      ground 3sg- warm
      "it is warm"

In both Juba Arabic and Kinubi *ras* 'head' also occurs in the complex preposition *fi ras* 'on':

(23)  a. Juba Arabic (**Nakao2012**)
      Merísa fí  fi  **ras**  terebéza
      beer   exs on head table
      "There is beer on the table."

(24)  b. Kinubi (**Wellens2003**)
      fi  **rás**  séder
      on head tree
      "on top of the tree"

**Nakao2012** attributes this function of *ras* to substratal influence from Acholi and Ma'di:

(25)  Acholi (**Nakao2012**)
      cib  **wi**-meja
      put head    table
      "Put it on the table!"

However, other possible sources include Western Nilotic languages such as Belanda Bor, Jur, Päri and Shilluk:


(26)  Jur (**PozzatiPanza1993**: 342)
      kedh ŋo  **wi**  tarabesa
      put  3sg head table
      "put it on the table"

Moreover, a preposition 'on' derived from the noun 'head' is also attested in Bongo (Central Sudanic) and Zande (Niger-Congo):

(27)  Bongo (Moi & al. 2014: 39)
      Ba  **do** mbaa
      3sg on car
      "He is on a car."

(28)  Zande (De **Angelis2002**: 288)

mo mai he  **ri**  ngua

2SG put 3SG on wood

"put it on the wood"

The verb *gal*/*gale*/*gali* 'say' is used in Juba Arabic and Ugandan Nubi as a complementizer, with *verba dicendi* and verbs of cognition:

(29)  a. Juba Arabic (**Miller2001**)

Úwo kélem **gal**    úwo bi=ǧa

3SG   speak COMP 3SG  IRR    come

"He said that he would come."

(30)  b. Ugandan Kinubi (**Wellens2003**)

úmon áruf **gal**  fí  difan-á  al  gi=ǧá

3PL    know COMP EXS guest-PL REL PROG  come

"they know that there are are guests who are coming"

The use of a *verbum dicendi* as a complementizer resembles the situation in Bari, where *adi* 'say' introduces direct speech (**Owens1997**: 163; **Miller2001**: 469) [5]:

(31)  Bari (**Miller2001**)

Mukungu a-  kulya **adi**    nan d'ad'ar kakitak merya-mukanat

sub-chief PST say    COMP 1SG want    worker fifty

"the sub-chief spoke saying: "I want fifty workers""

### 3.3.4 Lexicon

Since Bari is the main substrate language of Juba Arabic, unsurprisingly, it contributes most of its African-derived words: *gúgu* 'granary' < Bari *gugu*, *kení* 'co-wife' < Bari *köyini*, *loɲumég* < Bari *lónyumöng*, *tóŋga* 'pinch' < Bari *toŋga*. In several cases, the Juba Arabic form can be traced to a specific dialect: *d'oŋóŋ* 'back of head' < Pojulu *doŋoŋ*, *láŋa* 'wander' < Mundari *laŋa* 'travel', *ɲéte* vs *ŋéte* 'black-eyed pea leaf' < Bari *nyete* vs Kakwa, Pojulu *ŋete*. Moreover, "more Bari lexical items are being borrowed" in Youth Juba Arabic (**Nakao2012**): *kapaparát* 'butterfly' < Bari *kapoportat*, *lukulúli* 'bat' < Bari *lukululi*. Several other

---

[5]Unsurprisingly, in Juba Arabic "the use of adi in Bari [is] the most frequent [...] in particular among speakers of Bari origin" (**Miller2001**).

substrate and adstrate languages have contributed to the lexicon of Juba Arabic (**Nakao2012**; 2015): *adúngú* 'harp' < Acholi *adunu*; *b'ónǧo* 'pumpkin' < Bongo *b'onǧo*; *báfura* 'cassava' < Dinka *bafora* 'manioc, (sweet) cassava'; *káwu* 'cow-pea' < Ma'di *kau*; *malangí* < bottle' < Bangala/Lingala *molangi*; *kámba* 'belt' < Swahili *kamba*; *imbíró* 'palm tree' < Zande *mbíró*. Some 60 lexical items found in the earliest records of Ugandan Kinubi can be traced back to various substrate languages (**Avram2017b**): *lawoti* 'neighbours' < Acholi *lawoti* 'fellow, friend'; *korufu* 'leaf' < Bari *korofo/kɔrɔpɔ* 'leaves'; *lwar* 'abscess' < Dinka *luär* 'pain of a swelling'; *seri* 'fence' < Lugbara *seri* 'plant used for fencing'; *mukuta* 'key' < Päri *mukuta*.

The influence of Luganda and Swahili as adstrate languages is already documented in early Ugandan Kinubi (**Avram2017b**): Ugandan Kinubi *kibra/kibera* 'forest' < Luganda *e-kibira*, *nyinveza* 'fix' < Luganda *nyweza* 'make firm, hold firmly'; *dirisa* 'window' < Swahili *dirisha*; *kibanda* 'shed' < Swahili *kibanda* 'small shed'. The lexicon of modern Ugandan Kinubi is characterized by a large number of loanwords from Luganda and Swahili (**Wellens2003**; **Nakao2012**: 133–134), such as: *mé(é)mvu* 'banana' < Luganda *amaemvu* 'bananas', *ntulége* 'zebra' < Luganda *e-ntulege*; *karibísha* 'welcome' < Swahili *karibisha* 'welcome', *sangá/šangá* 'be surprised' < Sw *shangaa*. In some cases, these loanwords have replaced previously attested compounds consisting of Arabic-derived elements[6]: early Ugandan Kinubi *mária bitá murhúm* 'widow', lit. 'wife of the deceased' vs. modern Ugandan Kinubi *mamwándu* 'widow' < Luganda *nnamuwandu*. As for the lexicon of modern Kenyan Kinubi, it is strongly influenced by Swahili. **Luffin2004** lists some 170 loanwords from Swahili (out of approximately 1,400 words recorded), from a wide range of domains, e.g.: *barabára* 'highway' < Swahili *barabara*, *serikáli* 'government' < Swahili *serikali*, *tafaúti* 'difference' < Swahili *tafauti*, *úza* 'sell' < Swahili *ku-uza*. Swahili has also contributed several function words: *badáye* 'after' < Swahili *baadaye* 'afterwards', *íle* 'these' < Swahili *ile*, *na* 'and, with' < Swahili *na*. Kenyan Kinubi lexical items have occasionally undergone semantic shift or semantic extension under the influence of the meanings of their Swahili counterparts (**Luffin2014**): *destúr* 'tradition', cf. Swahili *desturi* 'tradition'; *fáham* 'to understand, to remember', cf. Swahili -*fahamu* 'to understand, to remember'

In some cases, the exact origin of loanwords found in Juba Arabic cannot be established: *búra* 'cat' < Acholi, Bongo, Dinka, Päri *bura*, Didinga *buura*; *daŋá* 'bow' < Bari, Jur *daŋ*, Didinga *d'anga*, Dinka *dhaŋ*; *pondú* 'cassava leaf' < Bangala, Kakwa, Lingala *pondu*, Pojulu *pöndu*. The same holds for a number of loanwords attested in early Ugandan Kinubi (**Avram2017b**): *bongo* 'cloth' < Acholi,

---

[6]See also Tosco & **Manfredi2013**.

Lendu, Lugbara, Zande *bongo*, Bari *boŋgo*; *godogodo* 'thin from illness' < Acholi, Avokaya, Bari, Baka, Lotuho, Moru, Zande *godogodo* 'very weak, thin, sick(ly)'; *mukungu* 'headman' < Acholi, *mukuŋu*, Bari *mʊkʊŋgʊ*, Lugbara *mukungu*, Luganda *o-mukungu* '(sub-) chief'. This is also true of several Kinubi words attested in more recent sources (**Wellens2003**; **Nakao2012**: 133–134): *júju* 'shrew' < Bari *juju*, Ma'di *juju*; *kingílo* 'rhinoceros' < Avokaya *kiŋgili*, Moru *kingile*. In some cases, the occurrence of alternative forms is due to their different SLs: *banǧa* 'debt' < Bari *banja*, Lugbara *banja*, Luganda *e-bbanja* vs. *banya* 'debt' < Acholi *banya*.

Under the influence of the substrate and adstrate languages, some Arabic-derived lexical items have undergone semantic extension, thereby becoming polysemous in Juba Arabic (**Nakao2012**), e.g. *gówi* 'hard; difficult', cf. Acholi *tek*, Bari *logo'*, Lotuho *gol*, Ma'di *okpo*, Swahili *kali*.

Juba Arabic "compensates its lexical gaps through the lexification of Arabic morphosyntactic sequences" (**ToscoManfredi2013**: 509). A case in point are Juba Arabic compounds, formed via juxtaposition or with their two members linked by the possessive particle *ta* (**Manfredi2014a**: 308–309). These include calques after several substrate languages (**Nakao2012**), e.g. *ída ta fíl* 'elephant trunk', cf. Acholi *ciŋ lyec*, Bari *könin lo tome*, Dinka *ciin akɔɔn*, Jur *ciŋ lyec*, Lotuho *naam tome*, Shilluk *bate lyec*, lit. 'arm (of) elephant'. Kinubi also exhibits a number of calques (**Nakao2012**; **Avram2017b**). Some of these compounds and phrases can be traced to several SLs, as in the following early Ugandan Kinubi examples (**Avram2017b**): *gata kalam* 'decide, judge', cf. Acholi *ŋɔlɔ kop* 'decide, give judgment', Bongo *ad'oci kudo*, Jur *ŋɔl lubo*, Päri *ŋondi lubo*, lit. 'cut word/speech'; Dinka *wèt tèm* 'decide, give the sentence', lit. 'word cut'; *jua bita ter* 'nest', cf. Acholi *ot winyo*, Bari *kadi-na-kwen*, Belanda Bor *kwɔt winy*, Shilluk *wot winyo*, Zande *dumô zirê*, lit. 'house (of) bird'. Other calques, presumably more recent ones, reflect the growing influence of Swahili on Kenyan Kinubi (**Luffin2014**): *bakán wáy* 'together', cf. Swahili *pamoja* 'together', lit. 'place one', *mára wáy wáy* 'seldom', cf. Swahili *mara moja moja* 'seldom', lit. 'time one one'.

To conclude, SL agentivity accounts for the small number of loanwords and calques recorded in the earliest stage (i.e. pidginization) of Juba Arabic and Kinubi. At a later stage (i.e. after nativization), the larger number of loanwords and calques is a result of borrowing under RL agentivity.

# 4 Arabic-lexifier pidgins in the Middle East

## 4.1 Current state and historical development

Several Arabic-lexifiers pidgins emerged in the Middle East. These include Romanian Pidgin Arabic, Pidgin Madam, Jordanian Pidgin Arabic, and Gulf Pidgin Arabic. The first three, can be classified as work force pidgins[7]. Gulf Pidgin Arabic also started out as work force pidgin (**Smart1990**), but it is now an interethnic contact language[8] (**Avram2014**).

Romanian Pidgin Arabic (**Avram2010**) was a short-lived pidgin, formerly used on Romanian well sites in Iraq, in locations in the vicinity of Ammara, Basra, Kut, Nassiriya, Rashdiya and Rumaila. Romanian Pidgin Arabic emerged after 1974, when Romanian well sites started operating in Iraq. Romanians typically made up two thirds of the oil crews, with Arabs making up the final third. The first Gulf War and the subsequent withdrawal of the Romanian oil rigs put an end to the use of Romanian Pidgin Arabic.

Immigration of Sri Lankan women to Arabic-speaking countries is reported to have started in 1976 (**Bizri2010**), but the large influx into Lebanon came later, in the early 1990s. Pidgin Madam is spoken in Lebanon by Sri Lankan female domestic workers and their Arab employers, mostly in the urban centres of the country.

Jordanian Pidgin Arabic (Al-**Salman2013**) is used in the city of Irbid, in the Ar-Ramtha district in the north of Jordan, in interactions between Jordanians and Southeast Asian migrant workers of various linguistic backgrounds. However, only Jordanian Pidgin Arabic as spoken by Bengalis is documented.

Gulf Pidgin Arabic is a blanket term designating the varieties of pidginized Arabic used in Saudi Arabia and the countries on the western coast of the Arab Gulf, i.e. Kuwait, the United Arab Emirates, Oman, Bahrain, and Qatar.

## 4.2 Contact languages

The main languages involved in the emergence of Romanian Pidgin Arabic are Romanian, Egyptian Arabic – spoken by immigrant workers, and Iraqi Arabic. A small minority of the participants in the language-contact situation had some knowledge of English.

The other pidginized varieties of Arabic in the Middle East share the charac-

---

[7]These are pidgins which "came into being in work situations" (**Bakker1995**).

[8]Which is "used not just for trade, but also in a wide variety of other domains" (**Bakker1995**).

teristic of having various Asian languages as their substrate[9]. For Pidgin Madam the main contact languages are Lebanese Arabic, as the lexifier language, and Sinhalese. Another language, with a much smaller contribution, is English. In the case of Jordanian Pidgin Arabic, the contact languages are mainly Jordanian Arabic and Bengali. The contribution of English is very limited. As for Gulf Pidgin Arabic, it emerged in a contact situation of striking complexity. On the one hand, Arabic, the lexifier language is represented by several dialects, which are not all subsumed under what is known as Gulf Arabic, in spite of what the name of the pidgin suggests. On the other hand, the number of languages spoken by the immigrant workers is staggering: for instance, in the United Arab Emirates the 200 nationalities and 150 ethnic groups speak some 150 languages. Adding to the complexity of the language-contact situation is the fact that these languages are typologically diverse. Last but not least, English also plays role in interethnic communication, particularly in the service sector.

## 4.3 Contact-induced changes

### 4.3.1 Phonology

The phonology of all the pidginized varieties of Arabic in the Middle East exhibits the outcomes of SL agentivity, which also accounts for the occurrence of considerable intra- and inter-speaker variation (**Avram2010**: 21–22; **Bizri2014**: 393; **Avram2017a**: 133).

Consider first Romanian Pidgin Arabic. The following are features characteristic of speakers with Romanian as L1. The phrayngeals are either replaced or deleted: *habib* 'friend' < IA/EA *ḥabīb*, *mufta* 'key' < IA/EA *muftāḥ*; *saa* 'hour' < IA/EA *sāʕa*. Plain consonants are substituted for pharyngealized ones: *halas* 'ready' < IA/EA *ḫalāṣ*. Both velar fricatives are replaced: *hamsa* 'five' < A *ḫamsa*; *šogol* 'work (N)' < IA *šuġ(u)l*. Geminate consonants are degeminated: *sita* 'six' < IA/EA *sitta*. There is no distinction between short and long vowels, either in lexical items of Arabic origin or in those from English: *lazim* 'must' < IA/EA *lāzim*; *slip* 'sleep' < E *sleep*. A feature typical of speakers with Egyptian or Iraqi Arabic as L1 is the substitution of /b/ for Romanian or English /p/ and /v/: *bibul* 'people, men' < English *people*; *gib* 'give, bring' < English *give*.

Consider next a number of selected features, generally typical of Pidgin Madam, Jordanian Pidgin Arabic, and Gulf Pidgin Arabic. Pharyngeals are either replaced – Pidgin Madam *hareb* 'war' < LA *ḥareb*; Jordanian Pidgin Arabic *bisallih* 'repair (v)' < JA *biṣalliḥ* 'PROG.3SG.M-repair'; Gulf Pidgin Arabic *aksan* 'best' < GA *aḥsan*,

---

[9]**Bizri2014** therefore suggests the cover term "Asian Migrant Arabic pidgins".

*ḥut* 'put' < GA *ḥut* 'put.ɪᴍᴘ.2sɢ.ᴍ' – or deleted: Pidgin Madam *ēki* 'cry' < LA *eḥki* 'cry.ɪᴍᴘ.2sɢ.ꜰ'; Jordanian Pidgin Arabic *arabi* 'Arabic' < JA *ʕarabī*; Gulf Pidgin Arabic *araf* 'know' < GA *ʕaraf*. The pharyngealized consonants are replaced by plain counterparts – Pidgin Madam s*arep* 'envelope' < LA *ẓaref*; Jordanian Pidgin Arabic *bandora* 'tomato' < JA *banḍōra*; Gulf Pidgin Arabic *halas* < GA *ḫalaṣ* – or are realized as retroflex ones: Pidgin Madam *ṭawīle* 'long' < LA *ṭawīle* 'long-ꜰ.sɢ'. The velar stops are substituted by velar stops or, less frequently, by /h/ : Pidgin Madam *sokon* 'warm' < LA *soḫon* 'warm.ᴍ.sɢ', *sogol* < LA *šəġəl* 'work'; Jordanian Pidgin Arabic *kamsa* 'five' < JA *ḫamsa*, *sukul* 'work (ɴ)' < JA *šuġl*, *zagīr* 'small' < JA *ṣaġīr*; Gulf Pidgin Arabic *kubus* 'bread' < GA *ḫubz*, *halas* 'finish (v)' < GA *ḫalaṣ*; *yistokol* 'work' < GA *yištuġul* '3sɢ.ᴍ-work', *šugl* 'work' < GA *šuġl*. Geminate consonants generally undergo degemination (Næss 2008: 36; **Avram2014**: 15): Jordanian Pidgin Arabic *sitin* 'sixty' < JA *sittīn*; Gulf Pidgin Arabic *sita* 'six' < GA *sitta*.

Moreover, consonants not found in the L1s of the users of Gulf Pidgin Arabic may also be replaced. For instance, Indonesian, Javanese, Sinhalese and Tagalog speakers may substitute /f/ for /p/: Pidgin Madam *palēpil* 'falafel' < LA *falēfil*; Jordanian Pidgin Arabic *pi* 'in' < JA *fī* Gulf Pidgin Arabic *napar* 'person' < GA *nafar*; Indonesian and Sinhalese speakers may realize /z/ as [s] or [ʤ]: Pidgin Madam *esa* 'if' < LA *eza*; Gulf Pidgin Arabic *sēn* ~ *ʤēn* 'good' < GA *zēn* (**Bizri2014**: 393; **Avram2017a**: 133). Bengali and Sinhalese speakers may replace /ʃ/ with /s/: Pidgin Madam *sū* 'what' < LA *šū*; Jordanian Pidgin Arabic *su* 'what' < JA *šū*.

Finally, although phonetically long vowels do occur, vowel length is not distinctive, as shown by the occurrence of variation, e.g. Gulf Pidgin Arabic *baden* ~ *badēn* 'then' < GA *baʕdēn*.

### 4.3.2 Syntax

There is relatively little that can be attributed to SL agentivity in the syntax of the Arabic-lexifier pidgins in the Middle East (**Almoaily2013**; Al-**Salman2013**; **Avram2014**; **Bizri2014**; **Avram2017a**; **Bakir2017**).

Since the substrate of these varieties, with the exception of Romanian Pidgin Arabic, consists of many SOV languages, e.g. Bengali, Hindi/Urdu, Malayalam, Punjabi, Persian, Sinhalese, Tamil, this word order is occasionally attested (**Avram2017a**: 133–134; **Bizri2014**: 403). For instance, direct objects may occur in pre-verbal position:

(32)  a. Pidgin Madam (**Bizri2010**)

misṭer kilot      sīli
mister underwear take off
"Mister takes off his underwear"

(33)    b. Gulf Pidgin Arabic (**Avram2017a**)
ana čiko   sūp
1SG child see
"I will see my children"

In attributive possession constructions the order of constituents is possessor–possessee:

(34)    a. Pidgin Madam (**Bizri2010**)
kullu māmā   benet
all     mother girl
"all mother's girls"

(35)    b. Gulf Pidgin Arabic (Næss 2008: 87)
ana jaud      bādēn ysīr Jakarta, stokol
1SG husband then   go   Jakarta work
"Then my husband went to work in Jakarta"

Adjectives generally precedes the nouns they modify:

(36)    Pidgin Madam (**Bizri2010**)
**bīr** bēt
big house
"a big house"

Similarly, adverbs precede the adjectives they modify:

(37)    a. Pidgin Madam (**Bizri2010**)
**ṭīr**    gūḍ
very good
"very good"

(38)    b. Gulf Pidgin Arabic (**Avram2014**)
**sem** **sem**   kalām SA
same speak
"they speak in the same way"

Occasional instances of postpositions are attested:

(39)  a. Pidgin Madam (**Bizri2010**)
      mister **mayik** masārē
      mister with    money
      "the money is with Mister"

(40)  b. Gulf Pidgin Arabic (**Avram2014**)
      zamal **fok**
      camel above
      "on top of the camel"

Interestingly, Pidgin Madam has a focalized negative copula, derived etymologically from English *no*:

(41)  Pidgin Madam (**Bizri2010**)
      māmā  Bīrūt  **no**
      mother Beirut NEG.FOC
      "It's not in Beirut that my mother is."

This resembles the Sinhalese negator *nemiyi*, which "is used only in focalized phrases" (**Bizri2010**):

(42)  Pidgin Madam (**Bizri2010**)
      bat  kāve mama **nemeyi**
      rice ate  1SG   NEG.FOC
      "It is not I who ate the rice."

### 4.3.3 Lexicon

Imposition under SL agentivity accounts for the fact that there are few instances of transfer of lexical items from the various SLs into the non-dominant RL (i.e. the pidgin).

The lexicon of Romanian Pidgin Arabic includes words of Romanian and English origin (**Avram2010**): *mašina* 'car' < Romanian *maşină*, *sonda* 'oil rig' < Romanian *sonda*; *spik* 'speak, say, tell' < English *speak*, *tumač* 'much, many' < English *too much*. Occasionally, non-Arabic words undergo semantic extension under the influence of phonetically similar Arabic words (**Avram2010**): *gib* 'give; bring' < E *give*, cf. EA *gib* 'bring.IMP.2SGM'.

The lexicon of all the other pidginized varieties of Arabic spoken in the Middle East includes loanwords from English: Pidgin Madam *ambasi* < English *embassy*; *go* < English *go*, *kam* < English *come*, *no gūḍ* 'bad' < English *no good*, *oké* < English *OK*; Jordanian Pidgin Arabic *bēbi* 'child' < English *baby*, *finiš* 'finish (v)' < English *finish*, *fisa* 'visa' < English *visa*; Gulf Pidgin Arabic *hazband* < English *husband*, *patient* 'patient (adj.)' < English *patient*. However, as noted by **Smart1990** about Gulf Pidgin Arabic, "it is difficult to say […] whether they are a true part of the pidgin" or rather nonce borrowings.

Given the extreme diversity of the substrate, it is not surprising that only a few words from the SLs have made it into the lexicon of Gulf Pidgin Arabic (**Avram2017a**: 134–135): *ača* 'fine' < Urdu *achā* 'good, very well', *ğaldi/ğeldi/ğeldi* < Urdu/Hindi *jaldī* 'quick'.

Jordanian Pidgin Arabic and Gulf Pidgin Arabic exhibits light verb constructions which may well be calques after Bengali – noun/adjective + *kara* 'make' and respectively Urdu/Hindi – noun/adjective + *karnā* 'make' and/or Persian – noun/ adjective + *kardan* 'make': Jordanian Pidgin Arabic *sawwi zadīd* 'renew', lit. 'make new'; Gulf Pidgin Arabic *sawwi suāl* 'ask', lit. 'make a question', *sawwi zalān* 'upset', lit. 'make angry'.

## 5 Conclusion

This chapter has shown that Arabic-lexifier contact languages emerged primarily through imposition under SL agentivity, in line with the typology of contact languages (**Winford2005**: 396; 2008: 128).

The effects of imposition are most obvious in the phonology, syntax and the syntax-semantics interface, and to a lesser extent in the morphology and the lexicon. In the phonology, SL agentivity induces the loss or replacement of certain phonemes, not found in the SLs. However, there are also instances of imposition in the sense of transfer from the SLs. As seen, in e.g. Turku and Bongor Arabic, some consonants occur only in loanwords from the substrate languages. The occurrence of such loanwords confirms the fact that imposition under SL agentivity may include transfer of lexical items into the RL. Borrowing under RL agentivity has generally played a far less significant role in the development of Arabic pidgins and creoles. As expected, it mostly involves transfer of lexical items; these may lead to the borrowing of certain consonant phonemes, as seen in e.g. Juba Arabic and Kinubi. Finally, borrowing has been shown to include transfer of morphological material as well.

A notable difference between Juba Arabic and Kinubi, on the one hand, and

the Arabic-lexifiers in the Middle East, on the other hand, resides in the relative weight of imposition under SL agentivity and borrowing under RL agentivity. As seen, Juba Arabic and Kinubi exhibit the effects of both imposition – in their earliest stage (i.e. pidginization) – and of borrowing – in their latest stage (i.e. nativization. In contrast, imposition is pervasive in the Arabic-lexifiers pidgins in the Middle East, given that these varieties have not undergone nativization.

There are still a number of issues awaiting resolution. For instance, the identification of the SLs is rendered difficult by their number and typological diversity. This difficulty is further compounded by the fact that some substrate languages are still under- researched. This is particularly the case of the substrate languages of Juba Arabic and Kinubi. Also, the distinction between substrate and adstrate languages is blurred (**Nakao2012**), particularly when varieties emerge and develop *in situ*, e.g. Juba Arabic. Further research also needs to consider the effects of the existence of a creole continuum in Juba Arabic as well as of bilingual and monolingual speakers of the language on the relative importance of restructuring, imposition and borrowing. The extent of restructuring and imposition, for instance, is presumably much greater in basilectal and L2 varieties as opposed to acrolectal and monolingual varieties of the language. The same holds for Bongor Arabic, which, as shown, appears to be undergoing depidginization. Last but not least, further investigations are necessary to establish whether Gulf Pidgin Arabic is evolving towards stabilization, possibly becoming closer to its lexifier, via borrowing of morphological material or is rather undergoing constant repidginization, essentially via imposition.

## Further reading

**Miller1993**, (**Nakao2012**), and (**Luffin2014**) illustrate in detail the /substratal and adstratal influence on Juba Arabic and Kinubi.

(Avram in press) analyzes the substratal input in the lexicon of Turku.

**Avram2017a** and **Bakir2017** discuss the various sources of Gulf Pidgin Arabic.

## Abbreviations

CA Chadian Arabic
   CLS class
   EA Egyptian Arabic
   GA Gulf Arabic
   IA Iraqi Arabic

JA Jordanian Arabic
LA Lebanese Arabic
N noun
PRT particle
RL recipient language
SA Sudanese Arabic
SL source language
V verb

```
%%move bib entries to  localbibliography.bib

@book{Almoaily2013,
address = {Newcastle},
author = {Almoaily, Mohammad},
note = {(Doctoral dissertation).},
publisher = {Newcastle University},
title = {Language variation in Gulf {Pidgin} {Arabic}},
year = {2013}
}




@book{Al-Salman2013,
address = {Irbid},
author = {Al-Salman, Ibrahim},
note = {(Master's dissertation).},
publisher = {Yarmouk University},
title = {Jordanian {Pidgin} {Arabic}},
year = {2013}
}
```

Avram, Andrei A. in press. Africanisms in Turku. In Miller, Catherine, Barontini
} \textit{2} \textit{\citealt{June2007}}. Aix-en-Provence: Institut de recherches

```
@misc{Avram2017,
author = {Avram, Andrei A.},
note = {(Paper presented at the Eleventh Creolistics Workshop "Assessing old assump
24 \citealt{March2017}).},
title = {Evidence of variation in early records of {Nubi}},
year = {2017b}
}
```

```
@incollection{Avram2017,
address = {Turin},
author = {Avram, Andrei A.},
booktitle = {\textit{Linguistic} \textit{studies} \textit{in} \textit{the} \textit
editor = {Bettega, Simone and Gasparini, Fabio},
pages = {131–151},
publisher = {Università di Torino},
title = {Sources of Gulf {Pidgin} {Arabic} features},
year = {2017a}
}
```

```
@article{Avram2010,
author = {Avram, Andrei A.},
journal = {\textit{Journal} \textit{of} \textit{Language} \textit{Contact}},
pages = {19–37},
title = {An outline of {Romanian Pidgin Arabic}},
volume = {3},
year = {2010}
}
```

```
@article{Avram2014,
author = {Avram, Andrei A.},
journal = {\textit{Lengua} \textit{y} \textit{Migracion}},
number = {2},
pages = {7–40},
title = {Immigrant workers and language formation: {{G}}ulf {Pidgin} {Arabic}},
volume = {6},
year = {2014}
}




@book{Bakir2017,
address = {GPA genesis},
author = {Bakir, Murtadha J.},
note = {(Paper presented at the Annual Summer Conference of the Society for Pidg
22 \citealt{June2017}).},
publisher = {The extent of substratal influence},
year = {2017}
}




@incollection{Bakker1995,
address = {Amsterdam},
author = {Bakker, Peter},
booktitle = {\textit{{Pidgins}} \textit{and} \textit{creoles.} \textit{An} \text
editor = {Arends, Jacques, Muysken, Pieter and Smith, Norval},
pages = {130–157},
publisher = {John Benjamins},
title = {{Pidgins}},
year = {1995}
}
```

```
@incollection{Bakker2008,
address = {Malden, MA},
author = {Bakker, Peter},
booktitle = {\textit{The} \textit{Handbook} \textit{of} \textit{{Pidgin}} \textit{a
editor = {Kouwenberg, Silvia and Singler, John Victor},
pages = {130--157},
publisher = {Wiley-Blackwell},
title = {{Pidgins} versus {Creoles} and Pidgincreoles},
year = {2008}
}




@book{Bizri2010,
address = {Paris},
author = {Bizri, Fida},
publisher = {Geuthner},
title = {\textit{{Pidgin}} \textit{Madame.} \textit{Une} \textit{grammaire} \texti
year = {2010}
}




@article{Bizri2014,
author = {Bizri, Fida},
journal = {\textit{Journal} \textit{of} \textit{Pidgin} \textit{and} \textit{Creol
number = {2},
pages = {385--409},
title = {Unity and diversity across {Asian} migrant {Arabic} pidgins in the Middle
volume = {29},
year = {2014}
}
```

```
@incollection{BurengVincent1986,
address = {Dordrecht},
author = {Bureng Vincent, George},
booktitle = {\textit{Current} \textit{approaches} \textit{to} \textit{{African}}
editor = {Dimmendaal, Gerrit J.},
pages = {71–78},
publisher = {Foris},
title = {Juba {Arabic} from a Bari perspective},
volume = {3},
year = {1986}
}


@book{Coetsem1988,
address = {Dordrecht},
author = {Coetsem, Frans van},
publisher = {Foris},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two}
year = {1988}
}


@article{Coetsem1995,
author = {Coetsem, Frans van},
journal = {\textit{Leuvense} \textit{Bijdragen}},
pages = {63–85},
title = {Outlining a model of the transmission phenomenon in language contact},
volume = {88},
year = {1995}
}
```

```
@book{Coetsem2000,
address = {Heidelberg},
author = {Coetsem, Frans van},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theory}
year = {2000}
}




@article{Coetsem2003,
author = {Coetsem, Frans van},
journal = {\textit{Leuvense} \textit{Bijdragen}},
note = {. doi: 10.2143/LB.92..},
pages = {27-99},
title = {Topics in contact linguistics},
urldate = {1.542034},
volume = {92},
year = {2003}
}




@misc{Cook1905,
author = {Cook, Albert R.},
note = {Ms.},
title = {\textit{Dinka} \textit{and} \textit{{Nubi}} \textit{Vocabulary}},
year = {1905}
}




@book{DeAngelis2002,
address = {Trieste},
```

*??*

```
author = {De Angelis, Pietro},
publisher = {Edizioni Università di Trieste},
title = {\textit{Vocabolario} \textit{Zande}},
year = {2002}
}




@book{Jenkins1909,
address = {Kampala},
author = {Jenkins, Edward V.},
publisher = {The Uganda Company},
title = {\textit{{English}} \textit{−} \textit{{Arabic}} \textit{vocabulary} \te
year = {1909}
}




@incollection{Lucas2015,
address = {London},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{{Routledge}} \textit{handbook} \textit{of} \te
editor = {Bowern, Claire and Evans, Bethwyn},
pages = {519–536},
publisher = {Routledge},
title = {Contact-induced language change},
year = {2015}
}




@book{Luffin2004,
address = {4, \textit{Annexes.} \textit{Corpus} \textit{des} \textit{enregistrem
author = {Luffin, Xavier},
note = {(Doctoral dissertation).},
```

```
publisher = {Université Libre de Bruxelles},
title = {\textit{Un} \textit{créole} \textit{arabe:} \textit{le} \textit{kinubi} \
year = {2004}
}




@book{Luffin2005,
address = {Munich},
author = {Luffin, Xavier},
publisher = {Lincom Europa},
title = {\textit{Un} \textit{créole} \textit{arabe:} \textit{le} \textit{kinubi} \
year = {2005}
}




@incollection{Luffin2013,
address = {Berlin {\textbullet} Boston},
author = {Luffin, Xavier},
booktitle = {\textit{{African}} \textit{{Arabic}:} \textit{Approaches} \textit{to}
editor = {Lafkioui, Mena},
pages = {161–183},
publisher = {De Gruyter Mouton},
title = {Some new information about {Bongor} {Arabic}},
year = {2013}
}




@article{Luffin2014,
author = {Luffin, Xavier},
journal = {\textit{Journal} \textit{of} \textit{Pidgin} \textit{and} \textit{Creol
note = {. doi: 10.1075/jpcl.29.2.04luf.},
number = {2},
```

pages = {299–318},
title = {The influence of {Swahili} on Kinubi},
volume = {29},
year = {2014}
}

@article{Manfredi2014,
author = {Manfredi, Stefano},
journal = {\textit{RiCognizioni.} \textit{Rivista} \textit{di} \textit{lingue,}
number = {1},
pages = {53–61},
title = {Qualche osservazione sull'espressione del plurale nominale nel creole a
volume = {2},
year = {2014b}
}

@incollection{Manfredi2014,
address = {Rome},
author = {Manfredi, Stefano},
booktitle = {\textit{L'etimologia.} \textit{{Atti}} \textit{del} \textit{{XXXV}}
editor = {Manco, Alberto and Silvestri, Domenico},
pages = {305–311},
publisher = {Il Calamo},
title = {Strategie di rilessificazione in Juba {Arabic}},
year = {2014a}
}

@book{Manfredi2017,
address = {Leuven},

```
author = {Manfredi, Stefano},
publisher = {Peeters},
title = {\textit{Árabi-Júba:} \textit{un} \textit{pidgin-créole} \textit{du} \text
year = {2017}
}




@incollection{Manfredi2018,
address = {London},
author = {Manfredi, Stefano},
booktitle = {\textit{The} \textit{{Routledge}} \textit{handbook} \textit{of} \text
editor = {Benmamoun, Elabbas and Bassiouney, Reem},
pages = {407–420},
publisher = {Routledge},
title = {{Arabic} as a contact language},
year = {2018}
}




@misc{Meldon1913,
author = {Meldon, J. A.},
note = {Ms.},
title = {\textit{{English}} \textit{–} \textit{{Arabic}} \textit{dictionary} \text
year = {1913}
}




@incollection{Miller1989,
address = {Hamburg},
author = {Miller, Catherine},
booktitle = {\textit{Proceedings} \textit{of} \textit{the} \textit{4th} \textit{Ni
editor = {Bender, Lionel},
```

*??*

pages = {1–10},
publisher = {Helmut Buske},
title = {Bari interference in Juba {Arabic}},
year = {1989}
}

@article{Miller1993,
author = {Miller, Catherine},
journal = {\textit{Matériaux} \textit{arabes} \textit{et} \textit{sudarabiques-G
pages = {137–174},
title = {Restructuration morpho-syntaxique en Juba-{Arabic} et Ki-{Nubi}: à prop
volume = {5},
year = {1993}
}

@incollection{Miller2001,
address = {Leuven},
author = {Miller, Catherine},
booktitle = {\textit{Leçons} \textit{d'Afrique:} \textit{Filiation,} \textit{rup
editor = {Nicolaï, Robert},
pages = {455–482},
publisher = {Peeters},
title = {Grammaticalisation du verbe 'dire' et subordination en Juba {Arabic}},
year = {2001}
}

@book{MoiMoi2014,
address = {Juba},
author = {Moi, Daniel R., Kuduku, Mario L. B., Michael, Mary M., John, Simon H.,

```
publisher = {SIL-South Sudan},
title = {\textit{Bongo} \textit{grammar} \textit{book}},
year = {2014}
}




@book{Muraz1926,
address = {\textit{Vocabulaire} \textit{du} \textit{patois} \textit{arabe} \textit
author = {Muraz, Gaston},
publisher = {Charles Lavauzelle},
year = {1926}
}




@book{Næss2008,
address = {Oslo},
author = {Næss, Unn Gyda},
note = {(Master's dissertation).},
publisher = {University of Oslo},
sortname = {Naess, Unn Gyda},
title = {"Gulf {Pidgin} {Arabic}": {{I}}ndividual strategies or a structured varie
year = {2008}
}




@incollection{Nakao2012,
address = {Tokyo},
author = {Nakao, Shuichiro},
booktitle = {\textit{Challenges} \textit{in} \textit{Nilotic} \textit{linguistics}
editor = {Hieda, Osamu},
pages = {127–149},
publisher = {Research Institute for Languages and Cultures of Asia and Africa},
```

*??*

```
title = {Revising the substratal/adstratal influence on {Arabic} creoles},
year = {2012}
}




@book{Nakao2015,
address = {Another story of Bangala},
author = {Nakao, Shuichiro},
note = {(Paper presented at the 8th World Congress of African Linguistics, Kyoto
24 \citealt{August2015}).},
publisher = {Language contact with South Sudanese Arabic Creole},
year = {2015}
}




@book{Nebel1979,
address = {Bologna},
author = {Nebel, Arthur},
publisher = {Editrice Missionaria Italiana},
title = {\textit{Dinka} \textit{--} \textit{{{English}}.} \textit{{{English}}} \t
} \textit{Dinka} \textit{dictionary}},
year = {1979}
}




@book{OwenOwen1915,
address = {Bukalasa},
author = {Owen, H. R. and Keane, G. J.},
publisher = {White Father's Printing Press},
title = {\textit{An} \textit{abbreviated} \textit{vocabulary} \textit{in} \texti
year = {1915}
}
```

```
@article{Owens1985,
author = {Owens, Jonathan},
journal = {\textit{Anthropological} \textit{Linguistics}},
number = {2},
pages = {229–271},
title = {The origins of {East} {African} {Nubi}},
volume = {27},
year = {1985}
}




@incollection{Owens1997,
address = {Amsterdam},
author = {Owens, Jonathan},
booktitle = {\textit{Contact} \textit{languages.} \textit{A} \textit{wider} \texti
editor = {Thomason, Sarah G.},
pages = {125–172},
publisher = {John Benjamins},
title = {{Arabic}-based pidgins and creoles},
year = {1997}
}




@book{PozzatiPozzati1993,
address = {Trieste},
author = {Pozzati, Aurelio and Panza, G.},
publisher = {Edizioni Università di Trieste},
title = {\textit{Dizionario} \textit{Giur}},
year = {1993}
}
```

```
@book{Smart1990,
address = {Pidginization in Gulf Arabic},
author = {Smart, Jack R.},
note = {\textit{Anthropological} \textit{Linguistics} 32(1–2). 83–119.},
publisher = {A first report},
year = {1990}
}
```

```
@incollection{ToscoTosco2013,
address = {Oxford},
author = {Tosco, Mauro and Manfredi, Stefano},
booktitle = {\textit{The} \textit{{Oxford}} \textit{handbook} \textit{of} \texti
editor = {Owens, Jonathan},
pages = {496–519},
publisher = {Oxford University Press},
title = {{Pidgins} and creoles},
year = {2013}
}
```

```
@article{ToscoTosco1993,
author = {Tosco, Mauro and Owens, Jonathan},
journal = {\textit{Sprache} \textit{und} \textit{Geschichte} \textit{in} \textit
pages = {177–267},
title = {{Turku}: {{A}} descriptive comparative study},
volume = {14},
year = {1993}
}
```

```
@book{Trudgill2011,
address = {Oxford},
author = {Trudgill, Peter},
publisher = {Oxford University Press},
title = {\textit{Sociolinguistic} \textit{typology.} \textit{Social} \textit{deter
year = {2011}
}




@book{Wellens2003,
address = {The Nubi Language of Uganda. Nijmegen},
author = {Wellens, Inneke},
note = {(Doctoral dissertation).},
publisher = {Catholic University of Nijmegen},
title = {An {Arabic} {Creole} in {Africa}},
year = {2003}
}




@article{Winford2005,
author = {Winford, Donald},
journal = {\textit{Diachronica}},
note = {. doi: 10.1075/dia.22.2.05win.},
number = {2},
pages = {373--427},
title = {Contact-induced changes: {{C}}lassifications and processes},
volume = {22},
year = {2005}
}
```

*??*

```
@article{Winford2008,
author = {Winford, Donald},
journal = {\textit{Journal} \textit{of} \textit{Language} \textit{Contact}},
note = {. doi: 10.1163/ 00000008.},
number = {1},
pages = {124–145},
title = {Processes of creole formation and related contact-induced language chan
urldate = {087952522},
volume = {2},
year = {2008}
}
```

# Chapter 11

# Contact-induced change in Northern Domari

??

??

> This present chapter provides an overview of the linguistic outcomes of contact between Arabic and Northern Domari. Northern Domari is a group of dialects spoken in Syria, Lebanon, Jordan and Turkey. It remained until very recently largely unexplored. This article presents unpublished first-hand linguistic data collected in Lebanon, Syria, Jordan and Turkey. It focuses on the Beirut/Damascus variety with references to the dialects spoken in Northern Syria and Southern Turkey.

## 1 Current state and historical development

Domari is an Indic language spoken by the Doms in various countries of the Middle-East. The Doms are historically itinerant communities who specialize in service economies. This occupational profile led the lay public to call them the Middle Eastern Gypsies. Common occupations are informal dentistry, metalwork, instrument crafting, entertainment and begging. Most claim Sunni Islam as their religion with various degrees of syncretic practices. Although most of them gave up their semi-nomadic lifestyle and settled in the periphery of urban centers, mobility is still a salient element in the daily lives of many Doms.

The ethnonym Dom is mostly unknown to non-Doms who refer to them with various appellations such as *nawar, qurbāṭ* or *qarač*. The Standard Arabic word *ġaǧar* for "Gypsy" is variably accepted by the Doms who mostly understand European Gypsies. All these appellations are exonyms and the only endonym found across all communities is *dōm*. Only the Gypsies of Egypt, it seems, use a reflex of *ġaǧar* to refer to themselves.

From the 19th century onwards, European travellers reported its existence in the shape of word-lists collected in the Caucasus, Iran, Iraq and the Levant (see

**Herin2012** for a discussion of these sources). The first full-length grammatical description of a dialect of Domari is **Macalister1914**, who described the dialect spoken in Palestine in the first years of the 20th century. At present, the language is known to be spoken in Palestine, Jordan, Lebanon, Syria and Turkey. No recent account can confirm that it is still spoken in Iraq and Iran. There are roughly two dialectal areas: Southern Domari spoken in Palestine and Jordan, and Northern Domari spoken in Lebanon, Syria and Southern Turkey. This geographical division is not clear-cut because I recorded speakers of Southern varieties in Lebanon and speakers of Northern dialects in Jordan. The main isogloss separating these two groups is the maintenance of a two-way gender system. Southern dialects maintained gender distinction whereas it mostly disappeared in the north: *gara* '(s)he went' in the North vs *gara* 'he went' and *garī* 'she went' in the South Northern Domari and Southern Domari are sufficiently different to posit an early split. Mutual intelligibility appears to be very limited. A case in point is kinship terminology which is largely divergent in both groups. Within Northern Domari, the Beirut/Damascus dialect stands out because of the glottal realization [ʔ] of etymological /q/ and the loss of the differential subject marker *-ən*.

No general statement can be made about language endangerment. Jerusalem Domari is reported to have only one fluent speaker left (Matras, this volume), but there still must be fluent speakers of Palestinian Domari in the Gaza Strip and the West Bank. Young fluent speakers of Southern dialects are easy to find in Jordan. As far as Northern Domari is concerned, the language has ceased to be transmitted to the young generation in Beirut but not in Damascus. In Northern Syria, intergenerational transmission is quite solid. The situation in Southern Turkey is according to some consultants more precarious but I personally witnessed quite a few children fully conversant with the language. In any case, bilingual Doms acquire both languages in early childhood, making both languages "dominant" in Van Coestem's term (**Coestem2000**).

Many Dom groups are also found in Eastern Anatolia. These groups shifted to Kurdish but maintained an in-group lexicon based on Domari, locally called Domani. According to what I could personally observe on the ground and what well-informed local actors reported to me, full-fledged Domari is not spoken beyond Urfa. East of Urfa, the shift to Kurdish is complete and even the in-group lexicon is only remembered by elderly individuals. There are no reliable figures about the number of speakers of Domari. The language has often been mistaken for a variety of Romani but this claim has no linguistic ground, except that they are both classified as Central Indo-Aryan Languages with a possible Dardic adstrate.

Contact languages

Besides a Central Indic core and a Dardic adstrate, the language exhibits various layers of influence. Easily identifiable sources of contact are Persian, Kurdish, Turkish and finally Arabic. This suggests, quite logically, that the ancestors of the Doms left the Indian subcontinent, and then travelled into Persian speaking lands, before reaching Kurdish and Turkish speaking areas, most probably Eastern Anatolia before venturing into Arab lands. It is striking to see that the Iranian and the Turkic elements in Domari are not uniform across Northern and Southern varieties, which suggests an early split in Eastern Anatolia between speakers of both groups. The impact of Arabic also is not uniform both across Southern and Northern Domari, but also within Northern Domari. What this means is that the validity of any discussion of the Arabic component in Domari is limited to the varieties considered.

The Beirut/Damascus dialect is undoubtedly the most arabized one within the Northern Group, pointing to an older settlement of the community in an Arabic environment. Bilingualism (Domari-Arabic) is general in Lebanon and Syria. Except maybe very young children who still did not acquire any other language, monolinguals in Domari are not to be found. In Southern Turkey around Gaziantep, many are trilingual Domari-Turkish-Kurdish. In the Hatay province, many speakers above forty are trilingual Domari,ArabicTurkish. The generations born in the eighties onwards did not acquire Arabic.

According to personal recollection from various consultants, the community used to spend the winter in Lebanon, and would go back to Damascus in the summer. This semi-nomadic way of life seems to have stopped when the civil war in Lebanon began. Although movements between Beirut and Damascus were always continuous, this phenomenon ceased to be seasonal. In Damascus, they settled in the area of Sayyida Zaynab, in the suburbs of the city, and in Beirut many of them settled in Sabra. Since the civil war started in Syria, virtually all the Damascus community moved to Lebanon and settled in refugee camps in the Bekaa Valley close to the Syrian border.

## 2 Contact-induced changes in Northern Domari

As noted above, Domari speakers in Lebanon and Syria are also fully proficient in Arabic, to the point that I have never encountered or heard of any monolingual adult. The Dom community, although largely endogamous and socially isolated, cannot afford monolingualism, primarily because of their peripatetic profile. As far as one can judge, their proficiency in Arabic is that of any monolingual na-

tive speaker of Arabic. Their pronunciation, however, is often not fully congruent with the local dialect spoken in the immediate vicinity of their settlements. This is, as usual, due to the variety of inputs and migration after acquisition. The Doms of Beirut for instance, do not speak Beirut Arabic and their speech is immediately perceived as Syrian by Lebanese because they don't raise /ā/. Raising of /ā/ towards [eː] is the hallmark of Lebanese Arabic in perceptual dialectology. Proficient speakers of Domari all exhibit a balanced bilingualism Arabic-Domari. On the whole, there is a general license to integrate any Arabic lexeme, even when a non-Arabic morpheme exists. Code-switching is also very common and there seems to be no conservative ideology about linguistic practices, leading to a very permissive environment for language mixing.

## 2.1 Phonology

All the Arabic segmental phonology made its way into Domari. Arabic stands out cross-linguistically because of a series of back consonants such as the pharyngeals /ḥ/ and /ʕ/, the post-velars /q/, /x/ and /ġ/ and a set of velarized consonants whose number varies from dialect to dialect. Typically, sedentary varieties in the Levant minimally exhibit contrast between /ḍ/ ~ /ẓ/, /ṭ/ and /ṣ/. In Domari, the pharyngeals /ḥ/ and /ʕ/ are commonly found in loans from Arabic: *ḥḍər h-* 'watch' (from Levantine Arabic *ḥiḍir* 'he watched'). The same goes for /ʕ/: *ʕam-mər kar-* "build" (from Arabic *ʕammar* "he built"). An oddity surfaces in the word for coffee, realized *ʔaḥwa* from Arabic *ʔahwe*. These pharyngeals are also common in Kurdish derived items such as *ḥazār* 'thousand', *moʕōri* 'ant' and also in the inherited (Indic) stock in *ʕaqqōr* 'nut'. Post-velar /q/, /x/ and /ġ/ are found in all the layers of the language: *qāla* 'black' (inherited), *qāpī* 'door' (Turkish), *xāl* 'uncle', *sāġ* 'alive' (Kurdish), *ġarīb* 'strange' (Arabic). The most striking innovation of the Beirut/Damascus dialect is the glottal realization [ʔ] of /q/: *ʔər* 'son' (ST *qər*), *ʔāyīš* 'food' (ST *qāyīš*). This innovation is of course contact-induced because it is commonly found in the Arabic dialects of both Damascus and Beirut and beyond.

Velarized consonants mostly surface in the Arabic derived stock as in *naḍḍəf kar-* 'clean' (< Arabic *naḍḍaf* 'he cleaned'), but also in pre-Arabic items: *ḍāwaṭ* 'wedding' (borrowed from Kurdish but ultimately from Arabic *daʕwa* 'invitation'), *pāṣ* 'at him' (< Old Indo-Aryan *pārśvá* 'side'). It is still unclear to what extent velarization in Domari continues Indo-Aryan retroflexion (**Matras2012**). Domari also kept a contrast between /p/ and /b/, not found in Arabic: *bīrōm* 'I feared' vs *pīrōm* 'I drank'.

As far as vowels are concerned, Levantine Arabic exhibits either a two-way

short vowel system (/a/ and /ə/) or a three-way system (/a/, /i/ and /u/). In Northern Domari, only the two short vowels /a/ and /ə/ are contrastive: *kərī* 'house' vs *karī* 'pot'. Such a paucity of contrastive short vowels is probably due to contact with Arabic varieties which exhibit a two-way system (/a/ vs /ə/), such as many Lebanese and Syrian dialects. Most Arabic dialects in the area have a five-way system of long vowels because of the monophthongization of /ay/ and /aw/: /ā/, /ī/, /ū/, /ē/ and /ō/. In addition to these long vowels, Domari displays another contrast between /ā/ and a back /ą/ (I.P.A. [ɑː]): *māsī* [maːsiː] 'meat' (< Old Indo-Aryan *māṁsá*) vs *mą̄s-ī* [mɑːsiː] 'month.PL' (< Old Indo-Aryan *mā́sa*).

Domari also preserved distinct suprasegmental features such as final syllable stress assignment. Arabic derived items are fully integrated into that pattern and bear final primary stress, whether common nouns or proper nouns: Domari [faˈdya] vs Arabic [ˈfaːdya] (proper noun *Fādya*). An interesting phenomenon is that Arabic epenthetic vowels in final syllable are reinterpreted as plain vowels and bear primary stress. Compare Domari [sˤaˈʕab] and Arabic [ˈsˤaʕəb]) 'difficult'; Domari [waˈdˤaʕ] and Arabic [ˈwadˤəʕ] 'situation'.

## 2.2 Morphology

Northern Domari did not borrow any derivational or inflectional morpheme from Arabic. This is of course due to the fact that Arabic morphology is mostly non-concatenative. Borrowed morphology mostly came from Kurdish and Turkish whose morpheme segmentation is much more transparent. These morphemes must have entered Domari when Kurdish and Turkish were contact languages of Domari. A case in point is the Kurdish diminutive *-ək* made its way into all the layers of the lexicon: *panč-ək* 'tail', *xar-ək* 'bone' (both Indic), *taxt-ək* 'wood', *qannīn-ək* 'bottle' (both derived from Arabic *taxt* 'bed' and *qannīne* 'bottle'). The dialects of Northern Syria and Southern Turkey also borrowed from Kurdish the comparative suffix *-tar*, the Turkish conditional marker *-sa* and the Turkish superlative marker *ān*. These constructions are not available in the dialect of Beirut/Damascus which relies entirely on Arabic derived material. Compare the translation of the Arabic sentence *inte aḥsan minn-i* 'you are better than me' into Sarāqib Domari (**??**) and Beirut Domari (**??**):

(1)  Sarāqib Domari
     tō  dēšōm  bxēz-tar       ištōre
     2SG 1SG.ABL good-COMP COP.2SG
     'You are better than me'

(2) Beirut Domari
   tō  aḥsan wēšōm  ištōr
   2sg better 1sg.abl cop.2sg 'You are better than me'

Sarāqib is located in Northern Syria and the dialect spoken by the Doms of Sarāqib is a good representative of Northern Syria and Southern Turkey Domari. Three differences are immediately apparent. The first is morphological whereby there are different forms for the ablative of the 1st person pronoun. The second difference is syntactic: in (**??**) the standard precedes the comparative (*aḥsan wēšōm*) and in (**??**) it follows it (*dēšōm bxēz-tar*).[1] The Beirut Domari syntax exhibits full congruence with the Arabic syntax. The third difference is lexical. Because Beirut Domari does not have at its disposal the morpheme *-tar*, speakers are obliged to draw on Arabic for the comparative. This phenomenon, labelled bilingual suppletion by Matras, is described at length for Jerusalem Domari (**Matras2012**).

Beirut Domari also relies entirely on Arabic material for the expression of time and date, as shown in (**??**). In Northern Syria, speakers favour the use of inherited numerals, as exemplified in (**??**).

(3) Beirut Domari
   mānane        mi-s-sāʕa      ʕašra la s-sāʕa      sabʕa tmāne ōtanta sa
   stay.ipfv.1pl from-det-hour ten    to def-hour seven eight   there all
   čāǧ-an-sa                 'We stay there with the all the kids from 10 to 7 or 8'
   children-obl.pl-com

(4) Sarāqib Domari
   ḥatta saʕat štār ēwar    mānde      ē         čōrt-ə-ma
   until hour four evening stay.prf.3sg dem.obl wasteland-obl-in

   'He stayed until 4pm in this wasteland'

Some speakers of Beirut Domari also extend the use of Arabic to higher numerals because, according to their own judgment, they have difficulties retrieving the pre-Arabic options. A look at their distribution reveals that the main parameter that triggers the use of Arabic items is not so much high numerals, but rather the complexity of the numeral. Compare for that matter (**??**) and (**??**). In (**??**), the speaker uses Arabic for the more complex numeral '95000' but uses Domari items for simpler 2000, 3000 and 4000.

---

[1] Comparative constructions typically involve two noun phrases. **Stassen2013** labels the object of comparison the "the compare NP" and the other "the standard NP".

(5) Beirut Domari

| pārda | abōs | ša??-āka | ši | xamse | u | tisʕīn | alf | dolar |
|-------|------|----------|-----|-------|-----|--------|-----|-------|
| buy.PFV.3SG | 3SG.BEN | flat-INDF | about | five | and | ninety | thousand | dollar |

'He bought a flat for her, about ninety-five thousand dollars'

(6) Beirut Domari

| načīš-a-ki | dī | ḥazār | trən | ḥazār | štār | ḥazār | dfaʕ |
|------------|-----|-------|------|-------|------|-------|------|
| dancing-OBL-ABL | two | thousand | three | thousand | four | thousand | pay |

| kaštand | dādōs | kē |
|---------|-------|-----|
| do.PROG.3PL | her.mother | BEN |

'They give two, three, four thousand (dollars) to her mother from dancing'

As noted above, it appears that the use of Arabic numerals is closely linked to language dominance. Speakers themselves are aware of it and when asked why they don't use Domari numerals, they justify it claiming a lack of proficiency. Looking at the distribution of inherited and Arabic numerals is therefore a good way to assess whether language attrition is incipient or not.

The impact of Arabic is also apparent in some morphological differences between the Beirut/Damascus variety and the dialects of Northern Syria. For instance, the verb *sək-* means 'learn'. The Beirut/Damascus dialect adds the passive suffix *-yā* / *-ī*. The corresponding verb in Arabic *tʕallam* is marked with the valency decreasing prefix *t-*. What the speakers of Beirut/Damascus Domari did is replicate the valency decreasing prefix *t-* of *t-ʕallam* with the Domari passive suffix *yā* / *-ī*: *skə-rd-ōm* (Northern Syria, learn-PFV-1SG) vs *sk-ī-r-ōm* (Beirut/Damascus, learn-PASS-PFV-1SG) 'I learnt'.

Unlike Southern Domari, Northern Domari does not normally transfer Arabic plurals. Speakers simply use the singular form and add the Domari plural marker *-ī(n)*: *azʕar-īn* 'thugs' instead of the Arabic plural *zuʕrān*. Arabic plurals do surface at times but only when they exhibit a high degree of independence within the lexicon. Examples are *ʔarāyb-ē-mā* (relatives-PL-1PL) 'our relatives', *ğīrān-ē-mā* (neighbors-PL-1PL) 'our neighbors', from Arabic *qarāyib* and *ğīrān*. Although these items have singular forms (respectively *qarīb* and *ğār*), they are arguably lexicalized plurals and independent entries in the Arabic lexicon.

## 2.3 Syntax

### 2.3.1 Constituent order

The impact of Arabic in the realm of syntax is not uniform across Domari dialects. Dialects of Northern Syria and Southern Turkey show a strong tendency to-

wards a head-final constituent order typology, both within the NP and the clause. This feature is areal so its presence in Domari may well be contact-induced. For descriptive purposes I will consider this parameter to be the original one. The canonical syntax of the NP is (demonstrative) (numeral) (adjective) (noun) noun. Complex NP's could only be retrieved through elicitation (examples 7 to 10) and hardly occur in spontaneous speech. Example (**??**) illustrates the canonical syntax, where all the modifiers appear to the left of the head. Speakers of Beirut Domari tend to dislocate some modifiers to the right of the head, converging towards the Arabic syntax, as in (**??**), (**??**) and (**??**).

(7)  Sarāqib Domari
ē     štār lāfty-ən-ki       dād-ō-sā
DEM four girl-OBL.PL-ABL mother-SG-3PL
'The mother of these four girls'

(8)  Beirut Domari
dād-ō-sā        štār lāfty-an-ki
mother-SG-3PL four girl-OBL.PL-ABL
'The mother of the four girls'

(9)  Beirut Domari
nām-ē-sā      ğəwr-an-ki        tərn-an-ki
name-PL-3PL woman-OBL.PL-ABL three-OBL.PL-OBL
'The names of the three girls'

(10)  Beirut Domari
dōm-an-sa       ēr-an-sa        štār-an-sa
dom-OBL.PL-PL DEM-OBL.PL-COM four-OBL.PL-COM
'with these four Doms'

In (**??**), the speaker also dislocates to the right the numeral *trən* 'three' which normally appears to the left and should be *trən ğəwr-an-ki nām-ē-sā* (three woman-OBL.PL-ABL name-PL-3PL). The numeral remains unmarked for case when it appears to the left of the head. When it is placed to the right, it agrees in case with the head. This is also the case with the demonstrative in (**??**). Here the normal order should be *ē štār dōm-an-sa* (DEM four Dom-OBL.PL-COM). The fact that speakers replicate case-marking on right-dislocated modifiers suggest that they feel the need to strengthen constituency in case of non-canonical syntax.

The influence of Arabic also surfaces in the Beirut/Damascus dialect in the syntax of the quantifier *sa* 'all'. It is normally located to the right of the head:

*ammat sa* 'all the people' (people all). In Beirut, *sa* consistently surfaces to the left, like the Arabic quantifier *kull*: *sa ammat* (Arabic *kull in-nās*).[2]

### 2.3.2 Internal object

Domari regularly replicates Arabic constructions and idioms but tend to do so re-cruiting inherited or pre-Arabic material, i.e. they do not borrow Arabic material. For instance, all dialects replicated the so-called Arabic internal object construc-tion, commonly used in Arabic as a predicate modifying construction. Consider for instance (**??**) in Jordanian Arabic, where the speaker narrows the scope of the predication using the verbal noun *ʕirəf* 'knowledge', derived from the *ʕirif* 'he knew', and modifies it with the adjective *ṭayyib* 'good'. In (**??**), the speaker used the deverbal derivation *kūš* from the root *kū-* 'throw' and coded it as an object, as evident from the accusative marking *kūš-əs*. It replicates the Arabic internal object construction as illustrated in (**??**).

(11)  Jordanian Arabic
      baʕrif-hum              ʕirəf        ṭayyib
      know.ɪᴘꜰᴠ.1ꜱɢ-ᴏʙꜰ.3ᴘʟ knowledge good

      'I know them well'

(12)  Sarāqib Domari
      dādōs       ibnḥarām  e  ē   kūš-əs
      his.mother son.of.illicit ᴄᴏᴘ ᴅᴇᴍ throwing-ᴀᴄᴄ
      ktōs-s-e
      throw.ᴘʀꜰ.3ꜱɢ-ᴏʙꜰ.3ꜱɢ-ᴘʀꜱ

      'His mother is heartless for having thrown (her baby) in such way'

### 2.3.3 Impersonal construction

Speakers also replicate the Arabic impersonal construction with the indefinite *il-wāḥad* by way of the inherited noun *mānəs* 'individual, people'. Example (**??**) illustrates the use of *il-wāḥad* in (Jordanian) Arabic. In (**??**), the sequence *gzare māns-as* corresponds to Arabic *biʕiḍḍ il-wāḥad* literally 'it bites one'. The fact that *māns-as* replicates *il-wāḥad* is also apparent in the accusative marking in Domari, which normally surfaces only with definite objects. The referent here is

---

[2]Arabic *kull* can also appear to the right as in *in-nās kull-ha ~ kull-hum* 'all the people' but this is a marked syntax.

by nature indefinite and non-referential, so accusative marking in Domari can only be explained by the presence of the definite article *il-* in Arabic *il-wāḥad*.

(13)  (Jordanian) Arabic
      kān            ʕēb            il-wāḥad yrūḥ            ʕala ʔutēl
      be.PFV.3.M.SG shameful DEF-one go.SBJV.3.M.SG to    hotel
      'One was ashamed to spend the night in a hotel'

(14)  Beirut Domari
      ašti  ši  hana lli  baḥr-a-ma e    gzare        māns-as
      EXIST too DEM  REL sea-OBL-IN cop bite.IPFV.3SG man-ACC
      'There is this thing in the sea, it bites you'

### 2.3.4 Auxiliaries

Probably the most striking difference between Southern and Northern Domari as far as the Arabic component is concerned is the absence of Arabic inflected material in the North. Only the dialect of Beirut/Damascus borrowed the auxiliaries *kān* and its imperfective form *bikūn*, *ṣār* and *xalli*.

(15)  Beirut Domari
      ṣār            ğahhəz lakand lāfty-a        kē bxēr
      become.PFV.3SG prepare do.SBJV.3PL girl-OBL BEN well
      'They prepare the girl well now (for the wedding)'

(16)  Beirut Domari
      xaḍra  kān          məğnār-a
      Khadra be.PFV.3M.SG breastfeed.IPFV.3SG-PST
      'Khadra was breastfeeding'

(17)  Beirut Domari
      āwande        bikūn        krēnde      mā kē kyāmōr
      come.IPFV.3PL be.IPFV.3SG do.PRF.3PL 1SG BEN something
      '(My kids) would come and they would have done something (naughty)'

In (**??**), the subject is in the 3PL but *ṣār* remains invariable, as the 3PL is *ṣāru*. In (**??**), the subject is feminine so if there was agreement one would expect *kā-nat*, not masculine *kān*. A further intriguing feature in (**??**) is the redundancy in past marking, first with *kān* and second with the past suffix *-a*, which in Northern Syria and Southern Turkey Domari suffices to mark past tense. The same

invariability is apparent in (??) where the 3PL of *bikūn* should be *bikūnu*. These auxiliaries have the same semantic load as in Arabic. The morpheme *ṣār* puts emphasis on the inception of the event, *kān* followed by the imperfective places the event in the past and gives it an iterative/habitual aspect and *bikūn* describes a possible state of affair not attested at the time of utterance. Arabic *ṣār*, *kān* and *bikūn* are absent in the dialects of Northern Syria and Southern Turkey. The only auxiliary that was replicated is *ṣār*. These dialects however, only replicated the structure, not the substance, i.e. they rely on inherited morphemes, as exemplified in (??). The speaker simply translates Arabic *ṣār* with the Domari equivalent *hra*, replicating the Arabic structure *ṣār* + subjunctive (see 'Calquing', Manfredi, this volume). A further difference is word order with the verb placed clause-finally in the subordinate.

(18)  Sarāqib Domari
      hər            wārsīndạ lwār
      become.PFV.3SG rain     hit.SBJV.3SG
      'It started raining'

As noted above, in these dialects the functions of Arabic *kān* is expressed by the inherited past suffix *-a*. The functions covered by Arabic *bikūn* however do not seem to be encoded in the grammar of these dialects.

In Levantine Arabic, the imperative form *xalli* 'let' of *xalla* 'he let' is often used to soften an order and allows the speaker to avoid using an imperative, flagging a suggestion or an invitation, as shown in (??):

(19)  Jordanian Arabic
      xalli ibn-ak   yrūḥ         la ǧ-ǧēš
      let   son-2SG  go.SBJV.3SG  to DEF-army
      'Let your son serve in the army'

This auxiliary was borrowed into Beirut/Damascus Domari with the exact same function, as illustrated in (20). In this case too, *xalli* remains invariable and does not surface as *xallī-(h)un* (let.IMP.2SG-3PL) as it would in Beirut/Damascus Arabic. Here again, the dialects of Northern Syria and Southern Turkey borrowed the structure, but not the substance, and use the inherited root *mək* 'let', as exemplified in (21).

(20)  Beirut/Damascus Domari

xalli ğānd       dfən lakrand-əs
let   go.IPFV.3PL bury do.SBJV.3PL-3SG

'Let them go and bury him'

(21)  Aleppo Domari
mək pāʋər       pāsōr
let   come.IPFV.3SG 2SG.AD

'Let him come to your place'

### 2.3.5  Negation

Only two Arabic negators made their way into the grammar of Domari: Damascus Arabic *mū* and the contrastive negative coordination markers *lā...walā* 'neither...nor'. Arabic *mū* is only available in the dialect of Beirut/Damascus. Its distribution and functions however only partially match those of Damascus Arabic. The primary function of *mū* in Damascus Arabic is to negate non-verbal predicates. This is not attested in Domari which only relies on inherited *nye*. First, it surfaces when negation has scope over non-clausal constituents, as shown in (22) and second, when the predicate is in a non-indicative mood (subjunctive, jussive and imperative) as in (23):

(22)  Beirut Domari
səff (h)ra        wāšya   mū wāšōm
side become.PFV.3SG 3PL.COM NEG 1SG.COM

'He took sides with them, not with me'

(23)  Beirut Domari
biğūz   mū māntyar     wāš     məṣrī
possible NEG stay.SBJV.3SG 3SG.COM money

'He might not have any money left'

The Arabic structure *lā...walā* is readily available everywhere but whereas it is the only option in Beirut/Damascus, it competes with the inherited structure *nə...nə* in Northern Syria and Southern Turkey. This clash interestingly led to an intermediary form *nə...walā* as shown in (24). The Domari syntax is also reminiscent of the Turkish possessive predication syntax with possessive marking on the noun and an existential morpheme.

(24)  Antioch Domari (Southern Turkey)

nə    lawrōs ašti    wala šaršōs    ašti
NEG its.tree EXIST NEG its.root EXIST

'It doesn't grow on a tree nor has it roots'

### 2.3.6 Complex sentences

Complex sentences include minimally coordinated clauses and subordinate clauses. The Arabic coordinator *w* 'and', *aw* 'or', *walla* 'or', *bass* and others all made their way into Domari. Originally, Domari seems to have distinguished clausal coordination from phrasal coordination, a not so frequent feature from a typological point of view. Nominal categories are coordinated with the Turkish derived morpheme *la* and clauses are coordinated with the Kurdish derived enclitic *ši*. The intrusion of Arabic *w* which in Arabic is used indifferently for both kinds of coordination led to the marginalization of the original system in Beirut/Damascus Domari, which now tends to favour the use of Arabic *w*.

(25)   Beirut Domari
       illi mangar          tōre        māṣṭ-a-ma        w   illi mangar
       REL want.IPFV.3SG put.IPFV.3SG yoghurt-OBL-IN and REL want.IPFV.3SG
       ʔār-s-e                        nāšif
       eat.IPFV.3SG-OBJ.3SG-PRS dry

       'Some eat it in yoghurt, some eat it dry'

As far as phrasal coordination is concerned, some alternation between Arabic *w* and Turkish derived *la* is still observed: *dōmwārī u ṭāṭwārī ~ dōm la ʕarabi* 'Domari and Arabic'.

Virtually all the conjunctions of subordination found in Domari are from Arabic. This includes the relativizer *illi*, the complementizer *inno* and potentially all the adverbial conjunctions found in Levantine Arabic: *lamma* 'when', *qabəl-mā* 'before', *baʕd-mā* 'after', *ʕa-bēn-mā* 'by the time' and many more. Pre-Arabic constructions are attested for relativization and conditional clauses but these only survive in the dialects of Northern Syria and Southern Turkey and tend to be replaced by Arabic material (except in the varieties spoken in Turkey). A case in point is conditional clauses. Arabic *iza* and *law* are available everywhere, even in Turkey, as shown in (??), recorded in Antioch. In this example, the speaker uses the Arabic morpheme *aza* (< *iza*) in the first sentence of the utterance, and no overt marking in the protasis, making parataxis a possible means to express condition. As far as counterfactual conditions are concerned, it appears that the dialect of Beirut/Damascus is fully congruent with Arabic in having borrowed

also the morpheme *kān* in both the protasis and the apodosis, as shown in (**??**). The dialects of Northern Syria and Southern Turkey exhibit a native strategy using subjunctive mood and past marking in the protasis and perfective and past marking in the apodosis. The two clauses are coordinated with the Kurdish derived enclitic *ši* (28).

(26) Antioch Domari
 aza kām karne qāne kām nə-karne nə-qāne
 if work do.IPFV.1PL eat.IPFV.1PL work NEG-do.IPFV.1PL NEG-eat.IPFV.1PL
 'If we work, we eat, (if) we don't work, we don't eat'

(27) Beirut Domari
 law kān nəčnār-sā bābōm kān abṣar
 if be.PFV.3SG make.dance.IPFV.3SG-OBJ.3PL my.father be.PFV.3SG I.ignore
 kaki (h)re
 what become.PRF.3SG
 'If my father had put them to dance, I don't know what would have happened'

(28) Sarāqib Domari
 aḷḷ-əs byātyənd-a nə-ktēnd-s-a ši
 God-ACC fear.SBJV.3PL-PST NEG-throw.PFV.3PL-OBJ.3SG-PST and
 'Had they feared God, they would not have thrown him'

## 2.4 Lexicon

### 2.4.1 Function words

Arabic prepositions do occur in Domari but these are mostly non-core prepositions such as *qabəl* 'before', *baʕad* 'after', *minšān* 'for', *ġēr* 'other'. Some made their way into Domari only recently and still alternate with pre-Arabic options such as the Iranian morpheme *war*, which tends to be replaced by Arabic *mitəl* 'as, like' especially in the dialect of Beirut/Damascus. Currently, *war* and *mitəl* are in a quasi-complementary distribution with *war* being used with full NPs and *mitəl* with pronouns, as shown below in (29) and (**??**):

(29) Beirut/Damascus Domari
 tō ʔrōm war ištōr
 you my.son like COP.2SG
 'You are like my son'

(30)   Beirut/Damascus Domari
       tāni    ʔər gēna mitl-ōs  kām   karre
       second son also  like-3sg work do.ipfv.3sg
       'My second son has the same job'

The Arabic core preposition *b-* 'in, with' occurs in Domari but it appears to be restricted to certain constructions and idioms such as *gāl b-gāl* 'discussion' (word in-word), *ārāt əb-dīs* 'night and day' (night in-day), *b-rəbʕ-āk* 'for a quarter of a pound' (with-quarter-indf). The preposition *min* 'from' also sporadically occurs in Beirut/Damascus Domari:

(31)   Beirut/Damascus Domari
       min  ši    šēš mās   gərsā   krōm      ḍāwaṭōs
       from about six month wedding do.pfv.1sg his.wedding
       'Some six months ago I married (my son)'

Domari also borrows high-frequency adverbs, fillers, connectors and all kinds of discourse structuring devices such as *masala* 'for instance', *abadan* 'at all, never', *yaʕni* 'I mean', *aywa* 'yes, so', *waḷḷa* 'I swear', *inno* (complementizer and discourse marker) and many more. One finds also common adverbial phrases such as *ṭūl in-nhār* 'all day long', *ṭūl il-waḷ* 'all the time', *ʕala ṭūl* 'immediately'. The very common Domari phrase *tīka tīka* 'slowly' replicates Arabic *šwayy əwayy* .

## 2.4.2 Content words

In Syria and Lebanon, Arabic is the de facto lexical reservoir of Domari so there is a general licence to integrate any element from Arabic if no pre-Arabic option exists. The issue is the replacement of pre-Arabic options with Arabic material. There is of course a certain amount of variation in lexical knowledge across speakers but it seems possible to differentiate several levels of replaceability. Some items have long been replaced by Arabic words and only a handful of speakers are able to retrieve them such as *lōrga* 'tomato' or *pīsənga* 'bulgur', replaced respectively by Arabic *bandōra* and *bərǧəl*. Other items tend to be replaced by Arabic equivalents but may still surface in the speech of some speakers, such as *čatīn* 'hard', *cirkī* 'bird', *alčāx* 'low' replaced by Arabic *ṣaʕab*, *ṭēr ~ ʕaṣfūr* and *wāṭi*. Some items seem stable but are sporadically replaced with Arabic derived items such as *drəs kar-* 'study' instead of inherited *sək-*. Finally, other items such as

*ǧawwəz h-* 'get married' and *ǧirsāwĩ ir* freely alternate. It appears therefore that every pre-Arabic item is somewhere on a continuum of replaceability from 'very unlikely' to 'completely disappeared'. To illustrate the variablility in replaceability judgment, I remember an elicitation session in Aleppo with a father and his son. One of the sentences contained the Arabic word *baṣal* ''onion'. The son simply translated the sentence with the Arabic word *baṣal* but the father strongly objected to this answer stating that the proper Domari word is *pīwāz*.

As noted above, Arabic nouns are integrated in their singular form, except in the case of lexicalized plurals. Adjectives are borrowed in their masculine form and never agree in gender, as shown in (**??**). Except the past copula *a*, all the words are Arabic. Two features however allow their identification as Domari. First, *ḥāla* is realized without raising (also stressed on the last syllable [ħaˈla]), unlike Levantine Arabic *ḥāle,* and second *taʕbān* does not agree in gender with *ḥāla* and surfaces in its masculine form instead of feminine *taʕbāne* as it would normally occur in Arabic.

(32)   Beirut/Damascus Domari
ʔabəl   ḥāla      taʕbān a
Before situation tired   COP.PST
'Before, the situation was bad'

Arabic verbs are easily integrated into Domari because Domari has a light verb strategy. Roughly, transitive verbs tend to be integrated with the light verb *kar-* 'do': *rabbī kar-* 'raise' from Arabic *rabba-yrabbi* 'raise'. Intransitive verbs are integrated with *h-* 'become': *ʕīš h-* 'live' from Arabic *ʕāš-yʕīš* 'live'. While all the verbs that are integrated with *kar-* are transitive, some verbs integrated with *h-* are not intransitive: *lməs (h)rōs-s-e* 'he has touched it' (touch become.PFV.3SG-3SG-PRS) from Arabic *lamas-yilmis* 'touch'. This seems to happen with transitive verbs that are lower on the transitivity scale, or at least perceived so. In the case of *lamas-yilmis*, its integration into Domari by way of the light verb *h-* suggests that speakers perceive it as less transitive. Formally, speakers isolate the imperfective stem of the verb, and apply a vocalism in /i/: *nsī kar-* 'forget' and *stannī kar-* 'wait', from the Arabic imperfective stems *nsa* 'forget' and *stanna* 'wait'.[3] An exception to this tendency occurs with the so-called hollow roots in Arabic whose imperfective stem is CūC. In this case, speakers simply extract the imperfective stem and leave it unchanged: *zūr h-* 'visit', *dūr h-* 'turn', *ʕūz h-* 'need' from the Arabic imperfective stems *zūr, dūr* and *ʕūz.*

---

[3]These verbs are only available in Beirut/Damascus, other dialects use respectively *ziwra kar-* and *akī kar-*.

Some English derived items were also recorded in the dialect of Beirut/Damascus such as *mōmari* 'memory card', *hambarga* 'hamburger' and more surprisingly *tō-manǧīre* 'Tom & Jerry' [to:mandʒiːˈre], expectedly stressed on the last syllable.

### 2.4.3 Speech sample

Probably the best way to capture how Arabic integrates into Domari is to consider a piece of spontaneous speech, reproduced below in (**??**). It is part of a recorded discussion I was having with a consultant in her mid-thirties in Beirut. It illustrates the level of endangerment of Beirut Domari. The consultant belongs to the last generation of fluent speakers. Her children did not acquire the language. According to what she reports, she was unable to speak to her children in their early childhood because her husband, who is a semi-speaker of Domari, prevented her from transmitting the language. Her daughter-in-law, aged 21 at that time, is also a fluent speaker of Domari because she grew up in Damascus, where language transmission was more solid than in Lebanon. Both of them use Domari in the household. Her son reacts negatively when he hears it even labels it *aǧnabi* 'foreign, non-Arabic'. Linguistically, the text illustrates some of the features discussed above.

(33) Beirut Domari

nā n-ǧib karre, pānǧī gāl karre, gāl karre dōm
no NEG-tongue do.IPFV.3SG 3SG word do.IPFV.3SG Dom 1SG.COM 1SG

wāšōm, mā gāl kame wāšī ʕādi **bass** əʕrōm
word do.IPFV.1SG 3SG.COM normal but my.son shout do.IPFV.3SG

ʔzīn karre wat, ftyare ma-gāl ka
3SG.SUP say.3SG NEG-word do.SBJV.2SG foreign NEG-understand

**aǧnabí**, nə-**fəmm** (h)ōme watōr, gāl karse
become.IPFV.3SG 2SG.SUP word do.IPFV.2PL Arabic-OBL-IN I.mean

ʕarabiy-a-ma, **yaʕni**, ma-gāl k(a) ēhānī,
NEG-word do.SBJV.2SG so because NEG-understand

**laʔanno** n-fəmm (h)ōre watī, **bass** mā l pānǧī ǧib
become.IPFV.3SG 3SG.SUP but 1SG and 3SG tongue do.IPFV.1PL length

kane **ṭūl** **il-waʔət** kəry-a-ma **yaʕni iza** mā l
DEF-time house-OBL-IN I.mean if 1SG and 3SG COP.1PL

pānǧi štēn kəry-a-ma, **ṭūl in-nhār** gāl kane
house-OBL-IN length DEF-day word do.IPFV.1PL Dom-OBL-IN I.mean

dōm-a-ma **yaʕni**, X, ʔrōm wāri, ʕəmrōs **wāḥad u** **ʕišrīn**
X my.son bride her.age one and twenty year bigger

**sane,** **akbar** ʔrōm-ki b-trən wars, **mū** **ʕādi,** **ʕādi** nye,
my.son-ABL with-three year NEG normal normal COP.NEG 1PL must

amīn **lāzim** lpāran **azġar** wēšōma, **bass** bxēz e **u**
take.SBJV.1PL smaller 1PL.ABL but good COP and humane COP

**ādami** e, **u** **maḥšūm** e, mā ēhāny-a
and respectful COP.1SG so-OBL heart-OBL BEN take.PFV.1SG-OBJ.3SG

xr-a kē pārdōm-əs ʔrōm kē, **u** **ğamāʕt**ēm kē,
my.son BEN and my.folks BEN learn.PFV.3SG here then

skīr(a) ēta **baʕdēn** skīra **mahná** **baʕdēn** kām
learn.PFV.3SG profession then work do.PFV.3SG year-INDF year-INDF

əkra wars-ā wars-ā nīm **makanīk baʕdēn** wəndrārda
half mechanic then fire.PFV.3SG and now NEG=work

**u** īsa nə-kām kištar wala kkyā wēsre kəry-a-ma
do.PROG.3SG nor thing sit.PRF.3SG house-OBL-IN

'No, (my son) doesn't speak (Domari), (my daughter-in-law) does, she speaks with me, I speak with her normally but my son shouts at her and tells her "don't speak foreign, I don't understand you, you all speak in Arabic, don't speak like this", because he doesn't understand her. But me and her we speak all the time in Domari, that is, if the both of us are in the house, all day long we speak in Domari. X, the bride of my son, she is 21 years old, three years older than my son, it's not usual, we (women) have to take someone older, but she is a good person, humane and respectful. That's why I took her for my son and my family. (My son) studied here (in the school). After that he went for vocational training and worked for a year a year and a half then he renounced. And now he doesn't do anything, he stays home.'

## 3 Conclusion

Multilingualism seems to have been a normal state of affair amongst the Doms for a very long time, probably since the genesis of the community. The reason for this is mostly because the sociolinguistics of Domari has in likelihood remained unchanged throughout the centuries: Domari is a community language whose use is restricted to in-group communication. Out-group interactions imply the use of the majority language. Due to the very nature of their occupational profile, peripatetic groups are forced to have frequent interactions with outsiders. This involves *de facto* high levels of bilingualism. Although it is hard to assess whether

the dominant language is the insider code or the outsider code, it makes sense to suspect that balanced bilingualism was the norm, as much in the past as in the present.

Van **Coetsem2000** uses the term "transfer" generically for any kind of contact-induced phenomenon. If the transfer is triggered by speakers who are dominant in the source language, he uses the term "imposition". If it originates from recipient language dominance, it is called "borrowing". **Lucas2015** further introduces two categories, the first of which he calls 'restructuring', defined as a "type of change...brought about by speakers for whom the changing language is an L2, but it does not involve transfer". He notes that for individuals who acquired two languages simultaneously (in early childhood), "the distinction between borrowing and imposition breaks down". In this case, both languages typically undergo 'convergence', that is the fourth category of contact-induced change. Because I posit balanced bilingualism Arabic-Domari as the norm, the question that needs to be answered is whether all the contact-induced changes happening in Domari are the product of convergence, or whether there are changes that can be attributed to Arabic dominance agentivity? Another problem is the sociolinguistic limits of the model. Speakers with two first languages are expected to initiate changes that target both languages. When languages exhibit unbalanced sociolinguistic statuses (minority versus majority), one wonders how changes originating from minority language agentivity can diffuse to the majority. Although it cannot be ruled out, it remains very unlikely. Consequently, convergence will always happen in the direction of the minority language. And this is indeed what is happening between Arabic and Domari: they become more and more similar at all levels, but only Domari is moving towards Arabic.

In the realm of phonology, it was shown that Domari has kept a distinct inventory from Arabic, although convergence with Arabic is almost complete for short vowels. A possible consonantal imposition is found in Beirut Domari where etymological /q/ is realised /ʔ/, as in neighbouring Arabic dialects. As far as morphology is concerned, eligible candidates for imposition are the Kurdish diminutive -ək, the Turkish conditional clitic *sa* and superlative *ān*. An evident cases of imposition is the phenomenon that seems the most sensitive to dominance: the so-called 'bilingual suppletion' (**Matras2012**). Bilingual suppletion in Northern Domari can be observed only in the dialect of Beirut/Damascus in the case of comparatives, and incipiently in the case of numerals. As far as syntax is concerned, cases of imposition are probably the transfer of Arabic auxiliaries and the negator *mū*. The transfer of utterance modifiers such as fillers, adverbs, conjunctions and virtually all discourse structuring devices are so prone to replication in

contact situations (**Matras1998**) that it is difficult to assess the source of agentivity. Other features discussed in this paper, such as constituent order, the internal object and the impersonal construction are clear instances of convergence.

As noted above, the main direction change in Domari is towards convergence with Arabic, as expected in case of absence of dominance. The Beirut/Damascus is the most convergent one amongst all the Northern dialects, which in itself suggests that Arabic-Domari bilingualism is older in that variety. The Arabic component in Domari is largely uneven cross-dialectally and no overall statement about its nature can be made. The general picture that arises is that the impact of Arabic gradually increases from north to south, with the dialects located in Northern Syria and Southern Turkey being the least arabized and the Southern dialects spoken in Palestine and Jordan being the most influenced by Arabic, and the dialect of Beirut/Damascus exhibiting an intermediary stage. It was also shown that the main difference between Northern and Southern Domari as far as Arabic is concerned is the reluctance in Northern Domari to transfer Arabic inflections and the general tendency to favour the transfer of structures without substance.

## Further reading

For a general account of the Arabic component in all the varieties of Domari documented so far, see **Herin2018**. The paper discusses the Arabic component in Southern and Northern dialects. This is the only paper that tackles extensively the issue of contact-induced change in Domari from a global perspective. For a description of the Domari dialect of Aleppo, readers can refer to **Herin2012**. **Herin2014** identifies the grammatical features that make Northern Domari a coherent dialectal group. **Herin2016** investigates the full extent of variation in Domari as a whole, drawing on data from both Northern and Southern Domari. Readers can refer to Matras (this volume, for a list of references about Jerusalem Domari).

## Abbreviations

In addition to the Leipzig Glossing Rules, the following glosses are added: AD (adessive), IN (inessive), SUP (superessive), COMP (comparative).

```
%%move bib entries to  localbibliography.bib
```

```
@book{Coetsem2000,
address = {19). Heidelberg},
author = {Coetsem, Frans van},
publisher = {Winter},
title = {\textit{A} \textit{general} \textit{and} \textit{unified} \textit{theory}
year = {2000}
}




@article{Herin2012,
author = {Herin, Bruno},
journal = {\textit{Linguistic} \textit{Discovery}},
note = {. doi:10.1349/PS1.1537-0852.A.412.},
number = {2},
pages = {1–52},
title = {The Domari Language of Aleppo ({Syria})},
volume = {10},
year = {2012}
}




@article{Herin2014,
author = {Herin, Bruno},
journal = {\textit{Zeitschrift} \textit{der} \textit{Deutschen} \textit{Morgenländi
note = {. doi:10.13173/zeitdeutmorggese.164.2.0407.},
number = {2},
pages = {407–450},
title = {The {Northern} Dialects of Domari},
volume = {164},
year = {2014}
}
```

```
@article{Herin2016,
author = {Herin, Bruno},
journal = {\textit{Mediterranean} \textit{Language} \textit{Review}},
note = {. doi:10.13173/medilangrevi.23.2016.0033.},
pages = {33--73},
title = {Elements of Domari Dialectology},
volume = {23},
year = {2016}
}




@incollection{Herin2018,
address = {Amsterdam: John Benjamins Publishing Company. doi},
author = {Herin, Bruno},
booktitle = {\textit{Studies} \textit{in} \textit{{Arabic}} \textit{Linguistics}
editor = {Stefano Manfredi and Mauro Tosco},
note = {1075/sal.6.02her. https://benjamins.com/catalog/sal.6.02her (10 December
pages = {20--36},
publisher = {10},
title = {The {Arabic} component in Domari},
volume = {6},
year = {2018}
}




@incollection{Lucas2015,
address = {Routledge. doi},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{{Routledge}} \textit{Handbook} \textit{of} \te
editor = {Claire Bowern and Evans Bethwyn},
note = {4324/94013.ch24.  (6 January, 2018).},
```

```
pages = {519–536},
publisher = {10},
title = {Contact-induced language change},
url = {https://www.taylorfrancis.com/books/9781315794013},
urldate = {78131579},
year = {2015}
}




@misc{Macalister1914,
author = {Macalister, Robert Alexander Stewart},
note = {Edinburgh University Press.},
title = {\textit{The} \textit{language} \textit{of} \textit{the} \textit{{Nawar}} \
year = {1914}
}




@article{Matras1998,
author = {Matras, Yaron},
journal = {\textit{Linguistics} 36(2). doi:10.1515/ling},
note = {(17 December, 2018).},
title = {Utterance modifiers and universals of grammatical borrowing},
url = {https://www.degruyter.com/view/j/ling.1998.36.issue-2/ling.1998.36.2.281/lir
volume = {1998.36.2.281},
year = {1998}
}




@book{Matras2002,
address = {Cambridge},
author = {Matras, Yaron},
note = {Press.},
```

*??*

```
publisher = {Cambridge Univ},
title = {\textit{Romani:} \textit{a} \textit{linguistic} \textit{introduction}},
year = {2002}
}




@misc{Matras2012,
author = {Matras, Yaron},
note = {Mouton de Gruyter.},
title = {\textit{A} \textit{Grammar} \textit{of} \textit{Domari}},
year = {2012}
}
```

Stassen, Leon. 2013. Comparative Constructions. In Matthew S. Dryer \& Martin Ha

# Chapter 12

# Arabic in the diaspora

??

??

This paper offers an overview of contact-induced change in diasporic Arabic. It provides a socio-historical description of the Arabic diaspora, followed by a sociolinguistic profile of Arabic-speaking diasporic communities. Language change is analyzed at the phonological, morphological, syntactic and lexical level, distinguishing between contact-induced change and internal developments caused by reduced input and weakened monitoring. In the course of the description, parallels are drawn between diasporic Arabic and other contemporary or extinct contact varieties, such as Arabic-based pidgins and Andalusi Arabic.

This paper did not follow the citation guidelines, hence the citation were not retrieved

## 1 Current state and historical development

The terms Arabic in the diaspora and Arabic as a minority language have been used to designate two distinct linguistic entities, namely Arabic *Sprachinseln* outside the Arabic-speaking world and Arabic in contemporary migration settings. The two situations correspond to the two major social processes that give rise to language contact: conquest and migration. In the former case, speakers of Arabic were isolated from the central area in which the Arabic language is spoken, exposed to a different dominant language and consequently underwent a slow process of language erosion (and eventually shift) usually spanning across several generations. This situation often gives rise to long periods of relatively stable bilingualism, where contact-induced change is more noticeable (Sankoff, 2001, p. 641). In migration contexts, on the contrary, language shift occurs at a faster pace, sometimes within the lifespan of the first generation and usually no later than the third (Canagarajah, 2008, p. 151).

This chapter analyzes contact-induced change in migration contexts. Arab migration to the West started in the late nineteenth century, with the first wave of migrants who left Greater Syria to settle in the United States and Latin America. The first migrants were mostly Christian unskilled workers, followed by more educated Lebanese, Palestinians, Yemenis and Iraqis after World War II. During the 1950s and 1960s, more migrants continued to settle in the US, while the unstable political situations in Palestine, Lebanon and Iraq resulted in a fourth wave in the 1970s and 1980s (Rouchdy, 1992a, pp. 17–18). Because of the events that took place during the last two decades and that resulted in a further destabilization of the entire Middle East, immigration toward the US has never stopped, even though recent American policies have considerably reduced the intake of refugees and immigrants. **In2016**, however, 84,995 refugees were resettled in the US, with two Arabic-speaking countries (Syria and Iraq) featuring among the top five states that make 70% of the total intake.[1]

Large-scale migration to Western Europe from Arabic-speaking countries started in the wake of the decolonization process during the 1960s and mainly involved speakers from North Africa (Morocco, Algeria and Tunisia). Following a common trend in labor migration, men arrived first, followed by their wives and children. **In1995**, a total of 1,110,545 Moroccans, 655,576 Algerians and 279,813 Tunisians lived in Europe, mostly in France, the Netherlands, Belgium, Germany and Italy (Boumans and De Ruiter, 2002, pp. 259–260). The socioeconomic profile of the first immigrants mainly consisted of unskilled laborers, usually with low education rates. After six decades from the first wave of immigration, however, most communities consist today of a first, second and third generation, while the political upheaval which started at the end of 2010 resulted in a new wave of young immigrants. Both old and new immigrants had to face the economic crisis that hit Europe in the early 1990s and, again, in 2007, with particularly harsh consequences for the immigrant population (Boumans and De Ruiter, 2002, p. 261).

The sociolinguistic profile of Arabic-speaking communities in the diaspora is quite diverse in different parts of the world and can be analyzed using the ethnolinguistic vitality framework, according to which status, demographics, and institutional support shape the vitality of a linguistic minority (Giles et al., 1977; Ehala, 2015). Arabic-speaking immigrants do not usually enjoy a particularly high status, while the level of institutional support is variable. The first waves of immigration to the US, for instance, had to face an environment that was generally hostile to foreign languages. The English-only movement actively worked to impose the

---

[1]Data come from the US Department of State. https://www.state.gov/j/prm/releases/factsheets/2017/266365.htm, accessed November 16, 2017.

exclusive employment of English in public places, while the immigrants themselves committed to learning and using English to integrate into mainstream American life. Only in the aftermath of 9/11 did American policymakers begin to reevaluate the importance of Arabic (and other heritage languages), considering it a resource for homeland security (Albirini, 2016, pp. 319–320). Other countries, such as the Netherlands, provided higher levels of formal institutional support, including Arabic in school curricula. These efforts did not achieve the desired goals, however, mostly because the great linguistic diversity of the Moroccan community living in the Netherlands cannot be adequately represented in the teaching curricula. Moroccans in the Netherlands, in fact, speak different Arabic dialects, alongside three main varieties of Berber, namely Tašelḥit, Tamazigt and Tarifit (Extra and De Ruiter, 1994, pp. 160–161). The voluntary Home Language Instruction, however, provides instruction in Modern Standard Arabic, even though writing skills are only taught starting from 3rd grade (Extra and De Ruiter, 1994, pp. 163–165). This is not, of course, the language students are exposed to at home, but attempts to introduce Moroccan dialect or Berber are generally opposed by parents, who value Classical Arabic for its religious and cultural relevance. Similar Home Language Instruction programs are found in most European countries, even though their implementation is sometimes carried out by local governments (in the Netherlands and Germany), private organizations (in Spain) or even by the governments of the origin country (in France) (Boumans and De Ruiter, 2002, pp. 264–265). The Italian town of Mazara del Vallo in Sicily represents an extreme case, since the members of the Tunisian community obtained from the Tunisian government the opening of a Tunisian school, where a complete Arabic curriculum is offered and Italian is not even taught as a second language. This school, opened in 1981, was the first choice for Tunisian families until the end of the 1990s, who hoped for a possible return to Tunisia. When it eventually became clear that this was unlikely to happen, enrollments consequently declined, which means that Arabic teaching is no longer available to the community in any form (D'Anna, 2017a, pp. 73–77). Issues of diglossia and language diversity, thus, undermine Home Language Instruction programs, which usually occupy a marginal role within school curricula.

Given the generally low status of, and insufficient institutional support for, Arabic-speaking communities in the diaspora, demographic factors are often decisive in determining the ethnolinguistic vitality of the community. While speakers of Arabic are usually scattered in large areas where the dominant language is prevalently spoken, in some Dutch towns Moroccan youth make up 50% of the population of certain neighborhoods (Boumans, 2004, p. 50). At the other end of

the continuum, we find closely-knit communities, living in the same neighborhood, such as in Mazara del Vallo, where Tunisians hailing from the two neighboring towns of Mahdia and Chebba constitute up to 70% of the population of the old town (D'Anna, 2017a, p. 27). All things equal, given the low status of the Tunisian community and the mediocre institutional support they receive, it is primarily demographic factors which have resulted in the preservation of Arabic in this community beyond the threshold of the third generation.[2]

In the light of what has been said above, and despite some notable exceptions, Arabic diasporic communities are characterized by relatively rapid processes of language shift, both in the US (Daher, 1992, p. 29) and in Europe (Boumans and De Ruiter, 2002, p. 282). This means that the processes of contact-induced change observed in diasporic communities of Arabic are generally the prelude to language loss. The importance of studying language change in migrant languages, however, resides in the fact that the same changes usually take place, at a much slower rate, in the standard spoken in the homeland. Internally motivated change in diasporic varieties, from this perspective, often represent an accelerated version of language change in the homeland. Contact-induced change, on the other hand, sometimes suggests parallels with the socially different process of pidginization (Gonzo and Saltarelli, 1983, pp. 194–195). The study of Arabic diasporic communities, thus, can help us shed light on the more general evolution of the language, with regard to both contact-induced and internally-motivated change.

## 2 Contact languages

Contact languages for diasporic Arabic-speaking communities include, but are not restricted to, American (Rouchdy, 1992b) and British English (Abu Haidar, 2012), Portuguese in Brazil (Versteegh, 2014, p. 292), French (Boumans and Caubet, 2000), Dutch (Boumans, 2007, 2004, 2000; Boumans and Caubet, 2000; Boumans and De Ruiter, 2002), Spanish (Vicente, 2005, 2008) and Italian (D'Anna, 2017a, 2018). Some contact situations are better described than others, as in the case of English, French and Dutch. At the other end of the continuum, research in the outcome of contact between Italian and Arabic is extremely recent, and data on Portuguese are scarce.

In the following sections, we will draw from the sources so far cited to describe the main phenomena of language change occurring in diasporic Arabic at the

---

[2]Other factors also played a minor role in the preservation of Arabic in Mazara del Vallo (D'Anna, 2017a, pp. 80–81).

phonological, morphological, syntactic and lexical level, highlighting possible parallels with comparable changes in other non-diasporic varieties of Arabic.

## 3 Contact-induced changes in diasporic Arabic

Despite the great variety of contact languages, it is possible to individuate a number of phenomena that predictably occur in diasporic Arabic-speaking communities. It is not always easy, however, to assess whether an individual phenomenon is due to contact or whether it is, on the contrary, the result of internal development (Romaine, 1989, p. 377):

While it seems clear that some types of changes are due to interference from the dominant language, and others may be attributable to sociological and other external pressures, there are some changes which are language-internal. The latter type is in accordance with a principle of regularization and code reduction which one might expect when the language is acquired in a weakly monitored sociolinguistic environment […] (Gonzo and Saltarelli, 1977, p. 177).

The concept of WEAKENED MONITORING, a situation in which a generally accepted standard and the reinforcement of correct norms are lacking, is an effective tool of analysis when investigating language change in diasporic communities (Gonzo and Saltarelli, 1977, 1983). In a situation of weakened monitoring, processes of language change that are occurring slowly in other varieties of the language can be sped up.

In the following sections, interference between languages will be referred to as TRANSFER, which occurs from the SOURCE LANGUAGE (SL) to the RECIPIENT LANGUAGE (RL). If the speaker is dominant in the SL, transfer is more specifically defined as IMPOSITION. If, on the contrary, the speaker is dominant in the RL, transfer is defined as BORROWING (Van Coetsem, 1988, 2000; Lucas, 2015). While the concept of linguistic DOMINANCE will be extensively used in this paper, one final *caveat* concerns the difficulty of individuating the dominant language (which may actually shift) in second-generation speakers. Lucas identifies a category of 2L1 speakers, who undergo the simultaneous acquisition of two distinct native languages (Lucas, 2015, p. 525). The linguistic trajectory of most second-generation speakers, however, usually involves two consecutive stages in which first the heritage and then the socially dominant language function as the dominant language. While the heritage language is almost exclusively spoken at home during early childhood, in fact, second-generation speakers gradually shift to the socially dominant language when they start school and consequently expand their social network.

### 3.1 Phonology

In the domain of phonology, diasporic varieties of Arabic generally go in the direction of the loss of marked phonemes (Versteegh, 2014, p. 293). The phonemes undergoing erosion are generally the emphatic and post-velar ones, while the loss is not usually systematic, featuring a great deal of inter and intra-individual variation. In non-diasporic communities, adults, peers and institutions provide corrective feedback to children during their process of language acquisition, while in immigrant communities, due to the weakened monitoring mentioned above, the chain of intergenerational transmission is less secure. Some phenomena of phonetic loss thus have a developmental origin, and are equally common in pidgins and dying languages (Romaine, 1989, pp. 372–373). Consider the following example:

(1)  Tunisian Arabic, Mazara del Vallo (D'Anna, 2017a, p. 85)

ʕala ḫāṭr-i     ʕarbi u  nnəžžəm  naʕrəf    aktər wāəd

on   thought-OBL.1SG Arab and can.IMPF.1SG know.IMPF.1SG more one

mia lingua    "Because I'm an Arab and I can know above all my

my  language

language."

The speaker in sample (1) realizes the voiced pharyngeal fricative /ʕ/, one of the phonemes that are usually lost, but then fails to realize its voiceless counterpart /ḥ/ in *wāəd* < *wāḥəd* 'one'.[3] They also occur, as noted above, in Arabic-based pidgins and creoles, such as Juba Arabic (Manfredi, 2017, pp. 17, 21) (Avram, this volume).

In the process of phonological erosion, therefore, contact languages seem to have a limited impact. If the dominant language does not feature, in its phonemic inventory, the phoneme that is being eroded, it fails to reinforce whatever input young bilingual speakers receive in the other L1 in the contexts of primary socialization. Reduced input and weakened monitoring, however, play a bigger role, allowing forms usually observed in the earliest stages of language acquisition by monolingual children to survive and spread. It is relatively common, for instance, to observe the presence of shortened or reduced forms, such as *qe* < *lqe*

---

[3]Similar phenomena of phonetic simplification occur in peripheral varieties of Arabic and *Sprachinseln*, such as Nigerian Arabic (Owens, 1993, pp. 19–20) (Owens, this volume), Cypriot Maronite Arabic (Versteegh, 2014, p. 280) (Walter, this volume), Uzbekistan Arabic (Versteegh, 2014, p. 285) and Maltese (Borg and Azzopardi-Alexander, 1997, p. 299) (Lucas & Čéplö, this volume). The single varieties here mentioned vary with regard to the phonological simplification they underwent.

'he found', *ḥal < nḥal* 'bees', *ləd < uləd* 'kid', which sometimes give rise to phenomena of compensation, such as in *uləd > ləd > lədda* 'kid' (Tunisian diasporic Arabic, Mazara del Vallo, Italy) (D'Anna, 2017a, p. 85). In diasporic communities, reduced forms are more easily allowed to survive and spread, occurring in the speech of teenagers, as in the examples reported here. Once again, the same phenomenon also occurs in pidgin and dying languages:

> In the case of dying and pidgin languages it may be that children have greater scope to act as norm-makers due to the fact that a great deal of variability exists among the adult community (Romaine, 1989, pp. 372–373).

In conclusion, the phonology of diasporic Arabic does not seem to be heavily influenced by borrowing from contact languages. The combined action of reduced input and weakened monitoring, on the other hand, is responsible for the unsystematic loss of marked phonemes and for the survival and spread of reduced forms.

## 3.2 Morphology

The complex mixture of concatenative and non-concatenative morphology in the domain of Arabic plural formation has been one of the main focuses of research in situations of language contact resulting from migration. Once again, borrowing from contact languages and independent developments occur side by side.

In Arabic, both concatenative and non-concatenative morphology contribute to plural formation. Concatenative morphology, which consists in attaching a suffix to the singular noun, yields the so-called sound plurals, that is, in spoken Arabic, the plural suffixes *-īn* and *-āt* respectively. It has been argued that sound feminine plural is the default plural form according to the morphological underspecification hypothesis, even though masculine is the default gender in all other domains of plural morphology (Albirini and Benmamoun, 2014, pp. 855–856). While sound masculine plural is specified for (+human), in fact, sound feminine plural has the semantic feature (±human). Non-concatenative, or broken, plurals require a higher cognitive load, since they involve the mapping of a vocalic template onto a consonantal root[4]. Sound feminine plurals are acquired by children by the age of three, while broken plurals involving geminate and defective roots are not mastered until beyond the age of six (Albirini and Benmamoun, 2014, pp. 857–858). After the age of five, however, heritage speakers

---

[4]The notion of root and pattern, which has long been at the core of the morphology of Arabic, has been recently been criticized (Ratcliffe, 2013), even though psycholinguistic studies seem to confirm the existence of the root in the mental lexicon of native speakers (Boudelaa, 2013).

of Arabic become increasingly exposed to their L2, which encroaches upon their acquisition of broken plurals. It has thus been convincingly demonstrated that heritage speakers display a better command of sound plurals and that, in the domain of broken plurals, some are more affected by language erosion than others (Albirini and Benmamoun, 2014, pp. 858–859). Across different varieties of diasporic Arabic, therefore, plural morphology displays both contact phenomena due to borrowing and internal developments that are akin to what might be called RESTRUCTURING, that is:

changes that a speaker makes to an L2 that are the result not of imposition but of interpreting the L2 input in a way that a child acquiring an L1 would not (Lucas, 2015, p. 525).[5]

Borrowing from the contact languages can take two forms. In rare cases, the suffix plural morpheme of the contact language is directly borrowed, as in the examples *ḫuli-s* 'sheep-PL', *ḥmar-s* 'donkeys' and *l-ʕud-s* 'the horses'[6] collected from one Moroccan informant in the Netherlands (Boumans and De Ruiter, 2002, p. 274). Sometimes, however, transfer works in a subtler way, which consists in the generalization of the sound masculine plural suffix *-īn*,[7] by analogy with the default form of the contact language, yielding *ḫul-in* 'sheep-PL', *ḥmār-in* 'donkeys', *ʕewd-in* 'horses' (Boumans and De Ruiter, 2002, p. 274). A study conducted by Albirini & Benmamoun [put citation here] shows that L2 learners of Arabic usually tend to overgeneralize the sound masculine plural, wrongly perceived as a default form, while heritage speakers more often resort to the Arabic-specific default, i.e. sound feminine plural (Albirini and Benmamoun, 2014, pp. 866–867). The cases of borrowing reported above, therefore, represent an idiosyncratic exception.

On the other hand, the non-optimal circumstances under which Arabic is learned in diasporic communities often result in overgeneralization processes that cannot be directly attributed to contact. One of them is, as noted above, the generalization of the sound feminine plural *-āt*. In the domain of broken plurals, moreover, not all patterns are equally distributed. The iambic pattern, consisting of a light syllable followed by one with two moras (CVCVVC), is the most common among Arabic broken plurals (Albirini and Benmamoun, 2014, p. 857). As a consequence, it is often generalized by heritage speakers of Levantine varieties (Syrian, Lebanese, Palestinian and Jordanian) living in the US, yielding forms

---

[5]In this case, of course, the speaker would not be re-interpreting an L2, but an L1 learned under reduced input conditions and subject to language erosion.

[6]The target form is here *ʕewd-an*, so that also vowel quality is not standard.

[7]The suffix for masculine plural *–īn* is realized with a short vowel in the diasporic Moroccan varieties that are being discussed.

like *fallāḥ* 'farmer' pl. *aflāḥ / fulūḥ* (target plural *fallāḥ-īn*), *šubbāk* 'window' pl. *šubūk* (target plural *šabābīk*), *ṭabbāx* 'cook' pl. *ṭabāʔix* (target plural *ṭabbāx-īn*) (Albirini and Benmamoun, 2014, p. 865).[8]

Borrowing does not involve plural morphemes only, but other classes as well. In Mazara del Vallo, for instance, young speakers often use the Sicilian diminutive morpheme *-eddru* with Arabic names, creating morphological hybrids of the kind illustrated in (2):

(2)   Tunisian Arabic, Mazara del Vallo (D'Anna, 2017a, p. 107)
      Grazie safwani-ceddruu[9]
      thanks Safwan-DIM
      "Thanks little Safwan."

This type of borrowing, quite widespread among young speakers, seems to replicate another instance of contact-induced change that occurred in an extinct variety of Arabic. Andalusi Arabic, in fact, borrowed from Romance the diminutive morpheme *-el* (e.g. *tarabilla* 'mill-clapper' < *ṭarab+wlla* 'little music'), incidentally etymologically cognate with the Sicilian *-eddru* (Latin *-ellum* > Sicilian *-eddru/-eddu*) (Institute of Islamic Studies of the University of Zaragoza, 2013, p. 60). The behavior of the young Tunisian speakers of Mazara del Vallo, who use these Sicilian diminutives in a playful mode, might represent the first stage of the same process that resulted in in the transfer of this morpheme into Andalusi Arabic (D'Anna, 2017a, p. 108).

While plurals represent one of the most common areas of change in diasporic Arabic, morpheme borrowing is a much rarer phenomenon, which probably occurs in situations of more pronounced bilingualism. The above two examples, however, provide a representative exemplification of the effect of language contact in the domain of morphology.

## 3.3 Syntax

Borrowing and restructuring also happen in the domain of syntax. As has been noted both for Moroccans in the Netherlands (de Ruiter, 1989, p. 99) and Tunisians

---

[8]The overgeneralization of some broken plural patterns indicates that the root and pattern system is still productive in heritage speakers, as opposed, for instance, to speakers of Arabic-based pidgins and creoles. Recent studies, however, have advanced the hypothesis that the iambic pattern involves operations below the level of the word, but without necessarily entailing the mapping of a template onto a consonantal root (Albirini et al., 2014, p. 112).

[9]The utterance appeared as a Facebook post in the timeline of one of my informants and was transcribed verbatim.

in Italy (personal research), second-generation speakers tend to use simpler clauses than monolingual speakers, namely main or subordinate clauses to which no other clause is attached, as evident from the following sample:

(3)   Tunisian Arabic, Mazara del Vallo (personal research)
m-basi        əl-uləyyəd   rqad              u    l-kaləb
from-after DEF-boy.DIM sleep.PRF.3SG.M and DEF-dog also and
zāda                        u    l-žrāna   xaržət    mən əl-wāḥəd ēh
DEF-frog exit.PRF.3SG.M from DEF-one eh bottle
dabbūsa


"Then the little boy slept and also the dog and the frog escaped from the… bottle."

Accordingly, they also display the effects of language erosion in establishing long-distance dependencies typical of more complex clauses (Albirini, 2016, p. 305).

Palestinian and Egyptian speakers born in the US have also been found to realize overt pronouns in sentences that opt for the pro-drop strategy in the speech of monolinguals, which is probably due to the influence of English (Albirini et al., 2011, p. 283). Preliminary observations on second-generation Tunisians in Italy, in fact, do not show the same phenomenon. Since Italian is, like Arabic, a pro-drop language, the use of overt pronouns in American diasporic Arabic can be considered as a case of syntactic borrowing or convergence (**Lucas2015**), depending on the speakers' degree of bilingualism.

The syntax of negation is another area in which language erosion triggers phenomena that seem to be happening, albeit at a slower rate, in non-diasporic communities. Egyptian speakers in the US, for instance, seem to overgeneralize the monopartite negatior *miš* / *muš* at the expense of the default discontinuous verbal negator *ma…-š*:

(4)   Egyptian Arabic in the US (Albirini and Benmamoun, 2015, p. 482)
huwwa miš rāḥ              l-kaftiria
he        NEG go.PRF.3SG.M to-cafeteria

"He didn't go to the cafeteria."

The sentence in (3) represents a deviation from the standard Cairene dialect spoken by monolinguals. In Egypt, however, the negative copula *miš* / *muš* represents a pragmatically marked possibility to negate the *b-* imperfect (Brustad,

2000, p. 302), while in Cairo it is now the standard negation for future tense (*miš ḥa-...*, contrasting with *mā-ḥa-...-š* in some areas of Upper Egypt (Brustad, 2000, p. 285)). More generally, therefore, *mīš / muš* is gaining ground at the expense of the discontinuous negation (Brustad, 2000, p. 285), so that what we observe in diasporic Egyptian Arabic might just be an accelerated instance of the same process.

Another major area of language change, documented in most diasporic languages, is the erosion of complex agreement systems (Gonzo and Saltarelli, 1983, p. 192). In diasporic Arabic, heritage speakers show relatively few problems with subject–verb agreement, but struggle with the subtleties of noun–adjective agreement (Albirini et al., 2013, p. 8). Although verb agreement features more cells to be filled in its paradigm, noun–adjective agreement is less straightforward. Plural nouns, in fact, can trigger adjective agreement in the sound or broken plural or in the feminine singular depending on factors involving humanness, individuation and the morphological shape of both the noun and the adjective, with marked dialectal variation (D'Anna, 2017b, pp. 103–104). Heritage speakers thus perform significantly better when default agreement in the masculine singular is required (Albirini et al., 2013, p. 8), but display evident signs of language erosion when more complex structures are involved:

(5)  Egyptian Arabic in the US (Albirini, 2014, p. 740)
     wi-kamān baḥibb      arūḥ          l-Detroit ʕašān ʕinda-ha
     and-also love.IMPF.1SG go.IMPF.SUBJ.1SG to-Detroit because at-OBL.3SG.F
     maṭāʕim      *mumtaz-īn
     restaurant.PL excellent-PL.M
     "And I also like to go to Detroit because it has excellent restaurants."

In (5), the speaker selects the sound masculine plural, while non-human plural nouns require either the broken plural or the feminine singular in Egyptian Arabic. Once again, language change in diasporic Arabic, where the language is learned under reduced input conditions, tends to replicate processes of language change that happened or are happening in the Arabic-speaking world. In the case of agreement, the standardization that the agreement system underwent in the transition from pre-Classical to Classical Arabic has been convincingly explained as emerging from the overgeneralization of frequent patterns by L2 learners (Belnap, 1999).

Finally, isolated cases show syntactic borrowing or convergence[10] at the level of word order, which is usually preserved in diasporic contexts:

---

[10]Once again, considering this phenomenon as syntactic borrowing or convergence depends on

(6)   Moroccan Arabic in the Netherlands (Boumans, 2001, p. 105)

u    ʕṭat          l-u         dyal-u        l-lḥem
and give.PRF.3SG.F to-OBL.3SG.M GEN-OBL.3SG.M- DEF-meat

"And she gave it [i.e. the dog] its meat."

Sample (6) illustrates an extreme case of word order change, in which the possessive *dyal-u* 'its' precedes the head. Overgeneralization of permissible (but sometimes pragmatically marked) word orders, however, occur much more frequently. Egyptian heritage speakers in the US, for instance, use SVO order 77.65% of the time, vs 52.64% for Egyptian native speakers (Albirini et al., 2011, pp. 280–281).

In situations of stable bilingualism, such as in some Arabic *Sprachinseln*, convergence with contact languages can result in permanent alterations to word order. In Buxari Arabic, for instance, transitive verbs feature a mandatory SOV word order, with optional resumptive pronoun after the verb. Cleft sentences such as the following one are quite common in all Arabic dialects:

(7)   Egyptian Arabic (Ratcliffe, 2005, p. 145)
il-fustān   gibt-u
DEF-dress get.PRF.1SG-3SG.M

'I got the dress.'

In Buxari Arabic, which has long been in contact with SOV languages (such as Persian and Tajik), this structure became the standard for transitive verbs, so that the resumptive pronoun can also be dropped, as in the following sample:

(8)   Buxari Arabic (Ratcliffe, 2005) p.144
fāt      ʕūd  xada
INDEF stick take.PRF.3SG.M

'He took a stick.'

## 3.4  Lexicon

In the domain of lexical borrowing, which has attracted considerable interest among scholars, the situation of bilingualism in diasporic contexts poses some methodological issues in the individuation of actual loanwords. The production of heritage speakers, in fact, is inevitably marked by frequent phenomena of

---

the speaker's language dominance, which is not clear from the source and is not easily ascertained in second-generation speakers, whose dominant language is often subject to shift.

codeswitching, which makes difficult to distinguish between nonce-borrowings (Poplack, 1980) and codeswitching. If we define lexical borrowing as "the diachronic process by which languages enhance their vocabulary" (Matras, 2009, p. 106), in fact, it is not clear which language is here enhancing its vocabulary, since diasporic varieties of Arabic are not discrete varieties and feature the highest degree of internal variability. A possible solution to this impasse consists in looking exclusively at the linguistic properties of the alleged loanword. In this vein, Adalar and Tagliamonte (1998, p. 156) have shown that, when foreign-origin nouns appear in contexts in which they are completely surrounded by the other language, they are treated like borrowings (in this case, nonce-borrowings) at the phonological, morphological and syntactic level. When, on the contrary, they appear in bilingual (or multilingual) utterances, they represent cases of codeswitching, patterning with the language of their etymology. The domain of lexical borrowing in diasporic varieties of Arabic, however, is an area that needs further research.

## 4 Conclusion

This chapter has offered an overview of the main phenomena of contact-induced change observed in Arabic diasporic communities, distinguishing them from internal developments due to reduced input and weakened monitoring. Diasporic communities rarely feature situations of stable bilingualism, so that language change usually corresponds to language attrition and is followed by the complete shift to the dominant language. The study of language change in diasporic communities, however, constitutes an interesting field of investigation, both in itself and for the insight it can give us into language change in monolingual communities. Change at the phonological, morphological and syntactic level finds parallels in comparable phenomena that have occurred in the history of Arabic (such as in the case of agreement) or that are occurring as we speak (such as in the case of the spread of the negator *miš* in Egyptian Arabic). Not by chance, similar phenomena also occur(red) in the Arabic-based pidgins of East Africa, such as Juba Arabic. Various scholars, in fact, have maintained that the mechanisms of change differ in the degree of intensity, but not in their intrinsic nature, from those operating in less extreme situations of contact (e.g. Miller, 2003, p. 8; Lucas, 2015, p. 528).

On the other hand, the analysis of contact phenomena in diasporic communities poses some methodological issues with regard to the categories of borrowing, imposition and convergence (Van Coetsem, 1988, 2000). These categories, in fact,

imply the possibility to define clearly the speaker's dominant language or, at least, to define him as a stable 2L1 speaker. This is rarely the case with heritage speakers, whose repertoires follow trajectories in which language dominance shifts, usually from the heritage language to the socially dominant one. This process is usually concomitant with the beginning of school education, but we lack theoretical and methodological tools to determine with accuracy the speaker's position on the trajectory.

Further avenues of research on this topic thus include a more rigorous investigation of emerging and shifting repertoires and the analysis of the complex relation between diasporic languages, pidginization and creolization, which has already been the object of a number of contributions (e.g. Gonzo and Saltarelli, 1983; Romaine, 1989).

## Further reading

Diasporic Arabic has been the object of several studies, but none of them specifically targets contact-induced change, although language erosion and codeswitching are usually analyzed in depth. **Rouchdy1992b** is the first description of Arabic in the US, followed ten years later by **Rouchdy2002**, which analyzes more broadly language contact and conflict, with a section devoted to Arabic in the diaspora. **Owens2000** collects essays on Arabic as a minority language, focusing on both *Spracheninseln* and diasporic Arabic, but introducing also historical and cross-ethnic perspectives.

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}
```

Abu Haidar, F., 2012. Arabic and English in Conflict: Iraqis in the UK, in: Rouc Variations on a Sociolinguistic Theme. Routledge - Curzon, London - New York, pp 297.

```
@article{Adalar1998,
author = {Adalar, N., Tagliamonte, S.},
journal = {Int. J. Biling},
pages = {139—159},
```

title = {Borrowed nouns, bilingual people: {{T}}he case of the "Londrali" in {Nort
volume = {2},
year = {1998}
}


@misc{Albirini2016,
author = {Albirini, A.},
note = {Diglossia, Variation, Codeswitching, Attitudes and Identity. Routledge, Ne
title = {Modern {Arabic} Sociolinguistics},
year = {2016}
}


@article{Albirini2014,
author = {Albirini, A.},
journal = {Int. J. Biling},
pages = {730–765},
title = {Toward understanding the variability in the language proficiencies of {Ar
volume = {18},
year = {2014}
}


@article{Albirini2015,
author = {Albirini, A., Benmamoun, E.},
journal = {Biling. Lang. Cogn},
pages = {470–489},
title = {Factors affecting the retention of sentential negation in heritage {Egypt
volume = {18},
year = {2015}
}

```
@article{Albirini2014,
author = {Albirini, A., Benmamoun, E.},
journal = {Mod. Lang. J},
pages = {854–872},
title = {Concatenative and Nonconcatenative Plural Formation in L1, L2, and Heri
volume = {98},
year = {2014}
}
```

```
@article{Albirini2013,
author = {Albirini, A., Benmamoun, E., Chakrani, B.},
journal = {Biling. Lang. Cogn},
pages = {1–18},
title = {Gender and Number Agreement in the Oral Production of {Arabic} Heritage
volume = {16},
year = {2013}
}
```

```
@article{Albirini2014,
author = {Albirini, A., Benmamoun, E., Montrul, S., Saadah, E.},
journal = {Linguist. Approaches Biling},
pages = {89–123},
title = {{Arabic} plurals and root and pattern morphology in Palestinian and {Eg
volume = {4},
year = {2014}
}
```

@article{Albirini2011,
author = {Albirini, A., Benmamoun, E., Saadah, E.},
journal = {Stud. Second Lang. Acquis},
pages = {273–303},
title = {Grammatical features of {Egyptian} and Palestinian {Arabic} heritage speak
volume = {33},
year = {2011}
}

@article{Belnap1999,
author = {Belnap, R. K.},
journal = {Folia Linguist},
pages = {169–186},
title = {A New Perspective on the History of {Arabic} Variation in Marking Agreemen
volume = {33},
year = {1999}
}

@misc{Borg1997,
author = {Borg, A. J., Azzopardi-Alexander, M.},
note = {Routledge, London ; New York.},
title = {Maltese, Descriptive grammars},
year = {1997}
}

@book{Boudelaa2013,
address = {Psycholinguistics, in},

*??*

author = {Boudelaa, S.},
note = {Oxford University Press, Oxford, pp. 369–392.},
publisher = {The Oxford Handbook of Arabic Linguistics},
year = {2013}
}

Boumans, L., 2007. The periphrastic bilingual verb construction as a marker of i
311.

Boumans, L., 2004. L'arabe marocaine de la generation ayant grandi aux Pays-Bas,
68.

@article{Boumans2001,
author = {Boumans, L.},
journal = {Lang. Linguist. Rev. Int. Linguist},
pages = {97–120},
title = {Moroccan {Arabic} and {Dutch}: {{L}}anguages of Moroccan Youth in the {
volume = {8},
year = {2001}
}

Boumans, L., 2000. Topic Pronouns in Monolingual Arabic and in Codeswitching, in
95.

Boumans, L., Caubet, D., 2000. Modelling intrasentential codeswitching: a compar
181.

Boumans, L., De Ruiter, J.J., 2002. Moroccan Arabic in the European Diaspora, in: Variations on a Sociolinguistic Theme. Routledge - Curzon, London - New York, pp. 285.

@book{Brustad2000,
address = {The Syntax of Spoken Arabic},
author = {Brustad, K.},
note = {Georgetown University Press, Georgetown.},
publisher = {A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects},
year = {2000}
}

@article{Canagarajah2008,
author = {Canagarajah, S. A.},
journal = {J. Socioling},
pages = {143–176},
title = {Language shift and the family: {{Q}}uestions from the Sri {Lankan} {Tamil}},
volume = {12},
year = {2008}
}

@book{Daher1992,
address = {A Lebanese Dialect in Cleveland: Language Attrition in Progress, in},
author = {Daher, N. Y.},
note = {Wayne State University Press, Detroit, pp. 25–36.},
publisher = {The Arabic Language in America},
year = {1992}

*??*

}

@book{D'Anna2018,
address = {Dialect contact in the Tunisian diaspora},
author = {D'Anna, L.},
note = {Perspect. Arab. Linguist. XXXI.},
publisher = {Chebba speakers in Mazara del Vallo (Sicily)},
sortname = {D'Anna, L.},
year = {2018}
}

@misc{D'Anna2017,
author = {D'Anna, L.},
note = {Profilo sociolinguistico della comunità tunisina di Mazara del Vallo, Bi
title = {Italiano, siciliano e arabo in contatto},
year = {2017a}
}

@article{D'Anna2017,
author = {D'Anna, L.},
journal = {Folia Orient},
pages = {101–123},
sortname = {D'Anna, L.},
title = {Agreement with plural controllers in Fezzānī {Arabic}},
volume = {54},
year = {2017b}
}

@book{deRuiter1989,
address = {Young Moroccans in the Netherlands},
author = {de Ruiter, J. J.},
note = {Rijksuniversiteit, Utrecht.},
publisher = {an Integral Approach to their Language Situation and Acquisition of Du
year = {1989}
}


@book{Ehala2015,
address = {Ethnolinguistic Vitality, in},
author = {Ehala, M.},
note = {John Wiley \& Sons, Inc. https://doi.org/10.1002/11463.wbielsi046},
publisher = {The International Encyclopedia of Language and Social Interaction},
urldate = {97811186},
year = {2015}
}


@article{Extra1994,
author = {Extra, G., De Ruiter, J. J.},
journal = {Indian J. Appl. Linguist. Spec. Issue Arab. Outs. Arab World},
pages = {151–176},
title = {The sociolinguistic status of the Moroccan community in the {Netherlands}
volume = {20},
year = {1994}
}


Giles, H., Bourhis, R.Y., Taylor, D.M., 1977. Towards a theory of language in ethn:

*??*

348.

Gonzo, S., Saltarelli, M., 1983. Pidginization and linguistic change in emigrant
198.

Gonzo, S., Saltarelli, M., 1977. Migrant Languages. Linguistic Change in Progres
187.

@misc{InstituteofIslamicStudiesoftheUniversityofZaragoza2013,
author = {Institute of Islamic Studies of the University of Zaragoza},
note = {Brill, Leiden.},
title = {Descriptive and Comparative Grammar of Andalusi {Arabic}},
year = {2013}
}

Lucas, C., 2015. Contact-induced language change, in: Bowern, C., Evans, B. (Eds
536.

@book{Manfredi2017,
address = {Arabi Juba},
author = {Manfredi, S.},
note = {Peeters, Leuven - Paris.},
publisher = {un pidgin-créole du Soudan du Sud},
year = {2017}
}

```
@misc{Matras2009,
author = {Matras, Y.},
note = {Cambridge University Press, Cambridge.},
title = {Language Contact},
year = {2009}
}




@article{Miller2003,
author = {Miller, C.},
journal = {Al-Lugha},
pages = {7–46},
title = {The relevance of {{Arabic}}-based {Pidgins}-{Creoles} for {{Arabic}} Lingu
volume = {4},
year = {2003}
}




@misc{Owens2000,
author = {Owens, J. (Ed.)},
note = {Mouton - de Gruyter, Berlin - New York.},
title = {{Arabic} as a Minority Language},
year = {2000}
}




@misc{Owens1993,
author = {Owens, J.},
note = {Harrassowitz Verlag, Wiesbaden.},
title = {A Grammar of {Nigerian} {Arabic}, Semitica Viva},
```

*??*


year = {1993}
}



@article{Poplack1980,
author = {Poplack, S.},
journal = {Linguistics},
pages = {581–618},
title = {Sometimes I'll start a sentence in {Spanish} y termino en español},
volume = {18},
year = {1980}
}



Ratcliffe, R., 2005. Bukhara Arabic: A Metatypized Dialect of Arabic in Central
151.



Ratcliffe, R.R., 2013. Morphology, in: Owens, J. (Ed.), The Oxford Handbook of A
92.



Romaine, S., 1989. Pidgins, Creoles, immigrant and dying languages, in: Dorian,
385.



@misc{Rouchdy2002,
author = {Rouchdy, A.},
note = {Routledge - Curzon, London - New York.},
title = {Language Contact and Language Conflict in {Arabic} – Variations on a So
year = {2002}

}

Rouchdy, A., 1992a. Introduction, in: Rouchdy, A. (Ed.), The Arabic Language in Am
21.

```
@misc{Rouchdy1992,
author = {Rouchdy, A.},
note = {Wayne State University Press, Detroit.},
title = {The {Arabic} Language in {America}}},
year = {1992b}
}
```

Sankoff, G., 2001. Linguistic Outcomes of Language Contact, in: Trudgill, P., Cham
668.

```
@misc{VanCoetsem2000,
author = {Van Coetsem, F.},
note = {Winter, Heidelberg.},
title = {A general and unified theory of the transmission process in language cont
year = {2000}
}
```

```
@misc{VanCoetsem1988,
author = {Van Coetsem, F.},
note = {Foris, Dordrecht.},
```

*??*

```
title = {Loan phonology and the two transfer types in language contact},
year = {1988}
}
```

```
@misc{Versteegh2014,
author = {Versteegh, K.},
note = {ed. Edinburgh University Press, Edinburgh.},
title = {The {Arabic} Language, Second Edition},
year = {2014}
}
```

Vicente, Á., 2008. Two cases of Moroccan Arabic in the diaspora, in: Miller, C.,
145.

```
@misc{Vicente2005,
author = {Vicente, Á},
note = {Une étude sociolinguistique de sa communnauté musulmane. L'Harmattan, Pa
title = {Ceuta une ville entre deux langues},
year = {2005}
}
```

# Chapter 13

# Ḥassāniya Arabic

??

??

> The area where Ḥassāniyya is spoken, located on the outskirts of the Arab world, is contiguous with those of several languages that do not belong to the Afro-Asiatic phylum. However, the greatest influence on the evolution of Ḥassāniyya has been its contact with Berber and Classical Arabic. Loanwords from those languages are distinguished by specific features that have enriched and developed the phonological and morphological system of Ḥassāniyya. In other respects, Ḥassāniyya and Zenaga are currently in a state of either parallel evolution or reciprocal exchanges.

## 1 Current state and historical development

### 1.1 Historical development of Ḥassāniyya

The arrival in Morocco of the Banī Maʕqil, travelling companions of the Banī Hilāl and Banī Sulaym, is dated to the thirteenth century. However, the gradual shift to the territories further south of one of their branches – that of the Banī Ḥassān, the origin of the name given to the dialect described here – began closer to the start of the subsequent century.

At that time, the Sahel region of West Africa was inhabited by different communities: on the one hand there were the "white" nomadic Berber-speaking tribes, on the other hand, the sedentary "black" communities.

During the following centuries, particularly during the seventeenth and eighteenth centuries, the sphere of Zenaga Berber gradually diminished, until it ceased to exist in the 1950s, other than in a few tribes in the southwest of Mauritania. At the same time, Ḥassāniyya Arabic became the language of the nomads of the west Saharan group, maintaining a remarkable unity (Taine-**Cheikh2016**; 2018a).

There is virtually no direct documentation of the region's linguistic history during these centuries. This absence of information itself suggests a very gradual transformation and an extended period of bilingualism.

Despite the lack of documentation of the transfer phenomenon, it seems highly likely that bilinguals played a very important role in the changes described in this chapter.

## 1.2 Current situation of Ḥassāniyya

The presence of significant Ḥassāniyya-speaking communities is recognised in six countries. With the exception of Senegal and especially of Niger, the regions occupied by these communities, more or less adjacent, are situated primarily in Mauritania, in the north, northeast and east of the country.

The greatest number of Ḥassāniyya-speakers (approximately 2.8 out of a total of four million) are found in Mauritania, where they constitute the majority of the population (approximately 75%). The Ḥassāniyya language tends to fulfil the role of the lingua franca without, however, having genuine official recognition beyond, or even equal to, that which it has acquired (often recently) in neighbouring countries.

# 2 Contact languages

## 2.1 Contact with other Arabic varieties

The islamization of the Ḥassāniyya-speaking population took place at an early date, and Ḥassāniyya has therefore had lengthy exposure to Classical Arabic. For many centuries this contact remained superficial, however, except for among the Marabout tribes, where proficiency in literary Arabic was quite widespread and in some cases almost total. The teaching of Islamic sciences in other places reached quite exceptional levels in certain *mḥāḍ̣əṛ* (a type of traditional desert university).[1] In the post-colonial era, the choice of Arabic as official language, and the widespread Arabisation of education, media and services, greatly increased the Ḥassāniyya-speaking population's contact with literary Arabic (including in its Modern Standard form), though perfect fluency was not achieved, even amongst the young and educated populations.

---

[1]These may be referred to as universities both in terms of the standard of teaching and the length of students' studies. They were, however, small-scale, local affairs, located either in nomadic encampments or in ancient caravan cities.

Excluding the limited influence of the Egyptian and Lebanese–Syrian dialects used by the media, the Arabic dialects with which ḤassāniyyaHassaniyya comes in to contact with most often today are those of the neighbouring countries (southern Moroccan and southern Algerian). Most recently Moroccan koiné Arabic has established a presence in the Western Sahara, since the region came under Moroccan administration.

## 2.2 Contact with Berber languages

Ḥassāniyya has always been in contact with Berber languages. Currently, speakers of Ḥassāniyya are primarily in contact with Tašlḥiyt (south Morocco), Tuareg (Malian Sahara and the Timbuktu region) and Zenaga (southwest Mauritania). In these areas, some speakers are bilingual in Ḥassāniyya and Berber.

In Mauritania, where Zenaga previously occupied a much larger area, Berber clearly appears as a substrate.

## 2.3 Contact with languages of the Sahel

Contacts between Ḥassāniyya speakers and the languages spoken in the Sahel have varied across regions and over time, but have left few clearly discernible traces on Ḥassāniyya.

The contact with Soninke is ancient (cf. the toponym Chinguetti < Soninke *sí-n-gèdé* 'horse well'), but the effects are hardly noticeable outside of the old cities of Mauritania. The contact with Songhay is both very old and still ongoing, but is limited to the eastern part of the region in which Ḥassāniyya is spoken (especially the region of Timbuktu).

The influence of Wolof, albeit marginal, has always been more substantial in southwestern Mauritania, especially among the Awlād Ban$^y$ūg of the Rosso region. It peaked in the years 1950–70, in connection with the immigration to Senegal of many Moors (e.g. *gord$^y$igen* 'homosexual', lit. 'man-woman'). In Mauritania, the influence of Wolof can still be heard in some areas of urban crafts (e.g. mechanics, electricity), but it is primarily a vehicle for borrowing from French.

Although Pulaar speakers constitute the second-largest linguistic community of Mauritania, contact between Ḥassāniyya and Pulaar is very limited, with the exception of some bilingual groups (especially among the Harratins) in the Senegal River Valley.

Certain communities (particularly among the Fulani) were traditionally known for their perfect mastery of Ḥassāniyya. As a result of migration into major cities and the aggressive Arabisation policy led by the authorities, Ḥassāniyya has

gained ground amongst all the non-Arabic speakers of Mauritania (especially in the big cities and amongst younger people), but this has come at the cost of a sometimes very negative attitude towards the language.

## 2.4  Contact with Indo-European Languages

Exposure to French has prevailed in all the countries of the region, the only exception being the Western Sahara, which, since the end of the 19th century and until 1975, was under Spanish occupation.

In Mauritania the French occupation came relatively late and was relatively insignificant. However, the influence of the colonisers' language continued well after the country proclaimed its independence in 1960. That said, it has tended to regress since the end of the 20th century (especially with the rise of Standard Arabic, e.g. *minəstr* has been replaced by *wazīr* 'minister'), whilst exposure to English has become somewhat more significant, at least in the better educated sections of the population.

# 3  Contact-induced changes in Ḥassāniyya

## 3.1  Phonology

### 3.1.1  Consonants

#### 3.1.1.1  /ḍ/

As in the other bedouin dialects, /ḍ̣/ is the normal equivalent of the ⬚ of Classical Arabic (e.g. *ḍ̣mər* 'to have an empty stomach' (CA *ḍamira*) and *ḍ̣ḥak* 'to laugh (CA *ḍaḥika*). Nonetheless, /ḍ/ is found in a number of lexemes in Ḥassāniyya.

The form [ḍ] sometimes occurs as a phonetic realization of /d/ simply due to contact with an emphatic consonant (compare *ṣḍam* 'to upset' and *ṣadma* 'annoyance', CA ṢDM). However, /ḍ/ generally appears in the lexemes borrowed from standard Arabic, either in all words of a root, or in a subset of them, for example: *staḥḍaṛ* 'to be in agony' and *ḥaḍari* 'urbanite' but *ḥḍ̣aṛ* 'to be present' et *maḥᵊḍ̣ṛa* 'Qur'anic school'. The opposition /ḍ/ vs. /ḍ̣/ can therefore distinguish a classical meaning from a dialectal meaning: compare *staḥḍaṛ* to *staḥḍ̣aṛ* 'to remember').

/ḍ/ is common in the vocabulary of the literate. The less educated speakers sometimes replace /ḍ/ with /ḍ̣/ (as in *qāḍ̣i* for *qāḍi* 'judge'), but the stop realization is stable in many lexemes, including in loanwords not related to religion, such as *ḍʕīv* 'weak'.

The presence of the same phoneme /ḍ/ in Berber might have facilitated the preservation of its counterpart in standard Arabic loans, even though in Zenaga /ḍ/ is often fricative (intervocalically). Moreover, the /ḍ/ of Berber is normally devoiced in word-final position in Ḥassāniyya, just as in other Maghrebi dialects, for example: *ṣayvaṭ* 'to say goodbye', from Berber FḌ 'to send'.

### 3.1.1.2 /ẓ/

/ẓ/ is one of is one of the two emphatic phonemes of proto-Berber. This emphatic sibilant sound regularly passes from the source language to the recipient language when Berber words are used in Ḥassāniyya. For example: *aẓẓ/āẓẓ* 'wild pearl millet' (Zenaga *īẓi*).

However, /ẓ/ is also present in lexemes of a different origin. Among Ḥassāniyya roots also attested in Classical Arabic, *z often becomes /ẓ/ in the environment of /ṛ/ (e.g. *ṛāẓ* 'to try', CA *rāza*; *ṛaẓẓa* 'lightning', CA *rizz*, *ẓəbṛa* 'anvil', CA *zubra*). Sometimes /ẓ/ appears in lexemes with a pejorative connotation, e.g. *ẓṛaṭ* 'fart; lie' (CA *ḍaraṭa*), *ẓagg* 'make droppings (birds)' (CA *zaqq*).

### 3.1.1.3 /q/

The normal equivalent to the ⵇ of Classical Arabic is the velar stop /g/, as in other bedouin dialects (e.g. *bagra* 'cow', CA *baqara*). However, /q/ is in no way rare.

First of all, /q/ appears, like /ḍ/, in a number of words borrowed from Classical Arabic by the literate: *ʕaqʔd* 'religious marriage contract'; *vəssaq* 'to pervert'. The opposition /g/ vs. /q/ can therefore produce two families of words such as *qibla* 'Qibla, direction of Mecca' and *gəbla* 'One of the cardinal directions (south, southwest or west, depending on the region)'. It can also create a distinction between the concrete meaning (with /g/) and the abstract meaning (with /q/): *θgāl* 'become heavy', *θqāl* 'become painful'.

Next, /q/ is present in several lexemes of non-Arabic origin, such as *bsaq* 'silo', (in Walata) *raqansak* 'decorative pattern', (in southwest Mauritania) *sərqəlla* 'Soninke people', (in Néma) *sasundaqa* 'circumcision ceremony', *mzawṛaq* 'very diluted (of tea)', *asanqās* 'pipe plunger', *sayqad* 'shouting in public' and (in the southeast) *šayqa* 'to move sideways'. These lexemes, often rare and very local in use, seem to be borrowed mostly from the languages of the Sahel region.[2]

Finally, /q/ is the outcome of *ɣ in cases of gemination, (ɣɣ > qq): compare *raqqad* 'to make porridge' to *rɣīda* 'a variety of porridge' (CA *raɣīda*). This corre-

---

[2]I am currently unable to specify the origin of these terms except that *bsaq* (attested in Zenaga) could be of Wolof origin.

lation, attested in Zenaga and more generally in Berber, can be attributed to the substrate.

Insofar as the contrast between /ɣ/ and /q/ is poorly established in Berber, the substrate could also explain the tendency, sometimes observed in the southwest, to velarize non-classical instances of /q/ (or at least instances not identified as classical): hence *ġandīr* 'candle' for *qandīr* < CA *qandīl* – this is despite the fact that the shift /ɣ/ > /ʔ/ is very common in Zenaga. However, the influence of Berber does not explain the systematic shift of /ɣ/ to /q/ throughout the eastern part of the Ḥassāniyya region (including Mali): thus eastern *qlab* 'defeat' for southwestern *ɣlab* (CA *ɣalaba*).[3]

### 3.1.1.4 Glottal stop

The glottal stop is one of the phonemes of Zenaga (its presence in the language is in fact a feature that is unique among Berber varieties), however it is not found in Ḥassāniyya, with the exception of words borrowed from standard Arabic, e.g. *tʔabbad* 'to live religiously', *danāʔa* 'baseness' and *taʔḫīr* 'postponement'. Very rarely the glottal stop is also maintained when it occurs at the end of a word as in *baṛṛaʔ* 'to declare innocent'.

### 3.1.1.5 Palatalized consonants

There are three palatalized consonants, two dental (/tʸ/ and /dʸ/) and a nasal /nʸ/. Unlike the phonemes discussed above, these are very rare in Ḥassāniyya, especially /nʸ/.

The palatalized consonants are also attested in certain neighbouring languages of the Sahel, as well as in Zenaga (but these are not phonemes of Common Berber). They are rather infrequent in the Zenaga lexicon, occurring especially in syntagmatic contexts (-*d+y-*, -*n+y-*) and in morphological derivation (formation of the passive by affixation of a geminate /tʸ/).

In Ḥassāniyya, the palatalised consonants appear particularly in words borrowed from Zenaga or languages of the Sahel. Interestingly, certain loanwords from Zenaga are of Arabic origin and constitute examples of phonological integration, as in *tʸfāɣa*, a given name and, in the plural, the name of a tribe < Zenaga *atʸfāɣa* 'marabout' < CA *al-faqīh*, and *ḫurūdʸ* 'leave (from Qur'anic school)' < Zenaga *ḫurūdʸ* < CA *ḫuruǧ* 'exit'.

---

[3]The regular passage from /ɣ/ to /q/ is a typical Bedouin trait, related to the voiced realization (/g/) of *q. It occurs especially in southern Algeria, in various dialects of the Chad–Sudanese area, and in some Eastern dialects (**Cantineau1960**).

One should also note the palatisation of /t/ in certain lexemes from particular semantic domains (such as the two verbs related to fighting *t*ʸ*bəl* 'to hit hard' and *kawt*ʸ*am* 'boxer'). This may suggest the choice of a palatalised consonant for its expressive value (and would then be a marginal case of phonosymbolism).

### 3.1.1.6 Labial and Labiovelars

The labiovelar consonants (/mʷ/, /bʷ/, /fʷ/ and /vʷ/ or /ṃ/, /ḅ/, /f̣/ and /ṿ/) are common in Ḥassāniyya, as they are in Zenaga. In both cases, they often come in tandem with a realization [u] of the phoneme /ə/.

This phenomenon may have originally arisen in Zenaga since the Ḥassāniyya of Mali (where it was most likely in contact with other languages) exhibits greater preservation of a [u] vowel sound and, at the same time, less pronounced labiovelarization of consonants.

The Ḥassāniyya of Mali also has a silent use of the phoneme /f/, where the Ḥassāniyya of Mauritania is characterized by the use of /v/ in its place (**Heath2004**; an observation that my own studies have confirmed). This phonetic trait does not come directly from Zenaga (in which /v/ exists but is very rare), however, it could be connected with the preference for voiced phonemes in Berber generally and in Zenaga in particular.

### 3.1.2 Syllabic structures

In Ḥassāniyya the syllabic structures derived from Arabic do not contain short vowels in word-internal open syllables, with the exception of particular cases such as certain nouns of action (*ḥašy > ḥaši* 'filling') and passive participles in *u* (*mudagdag* 'broken'). However, loanwords from literary Arabic and other languages (notably Berber and French) display short vowels quite systematically in this context: *abadan* 'never' and *ḥazīn* 'sad' (from standard Arabic); *tamāt* 'gum' (from Zenaga *taʔmað*); *taṃāta* 'tomato'. In fact, it may be noted that, unlike the majority of Berber varieties (particularly in the north), Zenaga has a relatively substantial number of lexical items with short vowels (including *ə*) in open syllables: *kaṛaḍ* 'three', *tuðuṃaʔn* 'a few drops of rain' *awayan* 'languages', *əgəðih* 'necklace made from plants.[4]

---

[4]It is precisely for this reason that, regarding the loss of the short vowels in open syllables, I deem the hypothesis of a parallel evolution of syllabic structures in Maghrebi Arabic and Berber to be more convincing than the frequently held alternative hypothesis of a one-way influence of the Berber substrate on the Arabic adstrate.

Furthermore, a long vowel *ā* occurs word-finally in loaned nouns which in standard Arabic end with *-āʔ*: *vidā/vidāy* 'ransom'. In other cases, underlyingly long word-final vowels are only pronounced long when non-final in a genitive construct.

## 3.2   Morphology

### 3.2.1   Nominal morphology

#### 3.2.1.1   Standard forms

Nouns and adjectives borrowed from standard Arabic may often be identified by the presence of a) open syllables with short vowels, e.g.: *vaḍalāt* 'rest of a meal', *ɣaḍab* 'anger', *vasād* 'alteration', *ḥtimāl* 'possibility', b) short vowels /i/ (less frequently /u/) in a closed syllable: *miḥrāb* 'mihrab', *muḥarrir* 'inspector; editor'.

Some syllables are only attested in loanwords, such as the the nominal pattern CVCC, where the pronunciation of the double coda necessitates the insertion of a supporting vowel, in which case the dialect takes on the form CCVC: compare *ʕaqᵊd* 'religious marriage' with *ʕqal* 'wisdom'.

The most characteristic loanword pattern, however, is that of *taḥrīr* 'liberation; verification (of an account)'. In Ḥassāniyya the equivalent of the pattern $taC_1C_2\bar{\imath}C_3$ is $t\partial C_1C_2\bar{a}C_3$. For the root ḤRR, this provides a verbal noun for other meanings of the verb *ḥarrar*: *taḥrār* 'whipping of wool (to untangle it); adding flour to make dumplings'. As for the form $taC_1aC_2C_2uC_3$, the /u/ is sometimes lengthened: *taḥammul* 'obligation', but *tavakkūr* 'contemplation'.

#### 3.2.1.2   Berber affixes

Nouns borrowed from Berber are characterised by the frequent presence of the vowels /a, ā, i, ī, u, ū/. These are of varying lengths, except that in a word-final closed syllable they are always long and stressed. Since these vowels appear in all types of syllables – open and closed – this results in much more varied syllabic patterns than in nouns of Arabic origin.

These loans are also characterized by the presence of affixes which, in the source language, are markers of gender and/or number: the prefix *a/ā-* or *i/ī-* for the masculine, to which the prefix *t-* is also added for the feminine or, more frequently (especially in the singular), a circumfix *t-…-t*: compare *iggīw/īggīw* 'griot' with the feminine form *tiggiwīt/tīggīwīt*. A suffix in *-(ə)n* characterizes the plurals of these loanwords which, moreover, differ from the singulars in terms

of their vocalic form: *iggāwən/īggāwən* 'griots', feminine *tiggawātən/tīggawātən*. The presence of these affixes generally precludes the presence of the definite article.

Though these affixes pass from the source language to the target language along with the stems, the syllabic and vocalic patterns of such loans are often particular to Ḥassāniyya: compare Ḥassāniyya *āršān,* plural *īršyūn/īršīwən* 'shallow pit' with Zenaga *aʔraš,* plural *aʔraššan (see Taine-**Cheikh1997a***).

Ḥassāniyya speakers whose mother tongue is Zenaga have most likely played a role in the transfer of these affixes and their affixation to nouns of all origins (including those of Arabic origin: a possible example being *tasūvra* 'large decorated leather bag for travelling', cf. *sāvər* 'to travel'). The forms that these speakers use can also be different from those used by other Ḥassāniyya speakers – especially if the latter have not been in contact with Berber speakers for a long time.

It is not proven that Berber speakers are the only ones to have created and imposed these forms which are more "Berberized" than authentically Berber. However, it may be noted that the gender of nouns borrowed from Berber is generally well preserved in Ḥassāniyya, even for the feminine nouns losing their final *-t*, other than in special cases such as the collective *tayšaṭ* 'thorny tree *Balanites Aegyptiaca*' with a final *-ṭ* (< Zenaga *tayšaḌ* for *tayšaḍt*).[5] In fact, this indicates a deep penetration of the meaning of these affixes and of Berber morphology in general (up to and including the incompatibility of these affixes with the definite article).

The borrowing of the formants *ən-* 'he of' and *tən-* 'she of' (quasi-equivalents of the Arabic-derived *bū-* and *ūṃ(ṃ)-*) is fairly widespread, in particular in the formation of proper nouns. It is also mostly in toponyms and anthroponyms that the diminutive form with prefix *aɣ-* and suffix *-t* is found, e.g. the toponym Agjoujt (< *aġ-žoʔž-t* 'small ditch').

### 3.2.2 Verbal Morphology

#### 3.2.2.1 The Derivation of *sa-*

The existence of verb forms with the prefix *sa-* is one of the unique characteristics of Ḥassāniyya (**Cohen1963**; Taine-**Cheikh2003**). There is nothing, however, to indicate that the prefix is an ancient Semitic feature that Ḥassāniyya has preserved since its earliest days. Instead, the regular correspondences between the three series of derived verb forms (causative~factitive vs. reflexive vs. passive) and the

---

[5]In Zenaga, non-intervocalic geminates are distinguished not by length, but rather by tension, and it is this that is indicated by the use of uppercase for the final *Ḍ.*

specialization of the morpheme *t* as a specific marker of reflexivity underlie the creation of causative~factitives with *sa-*. Neologisms with *sa-* generally appear when forms with the prefix *sta-* have a particular meaning: *staslaʕ* 'to get worse (an injury)' – *saslaʕ* 'to worsen (injury)'; *stabṛak* 'to seek blessings' – *sabṛak* 'to give a blessing'; *stagwa* 'to behave as a griot' – *sagwa* 'to make someone a griot'; *staqbal* 'to head towards the Qibla' – *saqbal* 'to turn an animal for slaughter in the direction of the Qibla'.

Furthermore, the influence of Berber has certainly played a role since the prefix *s(a)-* (or one of its variants) very regularly forms the causative~factitive structure in this branch of the Afro-Asiatic language family.

In Zenaga, the most frequent realization of this prefix is with a palato-alveolar shibilant, but a sibilant realization also occurs, particularly with the roots of Arabic origin. For example: Hass. *sādəb* (variant of *ddəb*) – Zen. *yassiʔðab* 'to train an animal (saddle)' < CA ʔDB (cf. *ʔaddaba* 'educate, carefully bring up'); Hass. *sasla* – Zen. *yassaslah* 'to let a hide soak to give it a consistency similar to a placenta' and Hass. *stasla* – Zen. *staslah* 'start to lose fur (of hides left to soak)' < CA SLY (cf. *salā* 'placenta').

Parallel to these examples where the Berber forms (at least those with the prefix *st(a)-*) are most likely themselves borrowed, we also find patterns with *sa-/ša-* which are incontestably of Berber origin: compare Ḥassāniyya *niyyər* 'to have a good sense of direction', *sanyar* 'to show the way', *stanyar* 'to know well how to orient oneself' and Tuareg *ener* 'to guide', *sener* 'to make s.o. guide s.o.'. Typically, however, when Ḥassāniyya borrows causative forms from Berber, it usually integrates the Berber prefix as part of the Ḥassāniyya root, making it the first radical of a quadriliteral root, e.g. Hass. *sadba* – Tuareg *sidou* 'to make s.o. leave in the afternoon' and Hass. *ssadba (< tsadba)* – Tuareg *adou* 'to leave in the afternoon'.

The parallelism between Arabic and Berber is not necessarily respected in all cases, but the forms with initial *s/š* are usually causative/factative in both cases. The only exception concerns certain Zenaga verbal forms which have become irregular upon contact with Ḥassāniyya: thus *yassəðbah* 'to leave in the afternoon' or *yišnar* 'to orient oneself' (a variant of *yinar*), of which the original causative value is now carried by a form with a double prefix (*ž+š*): *yažəšnar* 'to guide'.

### 3.2.2.2 The Derivation of *u-*

The existence of a passive verbal prefix *u-* for quadrilateral verbs and derived forms constitutes another unique feature of Ḥassāniyya. For example: *udagdag*, passive of *dagdag* 'to break'; *uṭabbab*, passive of *ṭabbab* 'to train (an animal)',

*udāɣa*, passive of *dāɣa* 'to cheat (in a game)'.

The development of passives with *u-* was most likely influenced by Classical Arabic, since here the passives of all verbal measures feature a /u/ in the first syllable in both the perfect and the imperfect, e.g. *fuʕila yufʕalu*, *fuʕʕila yufaʕʕalu* et *fūʕila yufāʕalu*, the respective passives of *faʕa/i/ula*, *faʕʕala* et *fāʕala*.

However, influence from Berber cannot be excluded here since, in Zenaga, the formation of passives with the prefix *T$^y$-* is directly parallel to those of the passives with *u-* in Ḥassāniyya. Moreover, this prefix is *t(t)u-* or *t(t)w-* in other Berber varieties (especially those of Morocco) and this could also have had an influence on the emergence of the prefix *u-*.

## 3.3 Syntax

### 3.3.1 Ḥassāniyya–Zenaga parallelisms

Ḥassāniyya and Zenaga have numerous common features, and this especially true in the realm of syntax. In general, the reason for these common traits is that they both belong to the Afro-Asiatic family and remain conservative in various respects, for example, in their lack of a discontinuous negative construction.

There are, however, also features of several varieties of both languages documented in Mauritania that represent parallel innovations. Thus, corresponding to the diminutive forms particular to Zenaga, we have in Ḥassāniyya *mutatis mutandis* a remarkably similar extension of the diminutive pattern with infix -*ay-*, which is even applied to verbs, e.g. *mayllas*, diminutive of *mallas* 'to smooth over' (see Taine-**Cheikh2008a**: 123–124).

In the case of aspectual-temporal forms, there are frequent parallels, such as Ḥassāniyya *mā tla* and Zenaga *war yiššiy* 'no longer', Ḥassāniyya *ma-zāl* and Zenaga *yaššiy* 'still', Ḥassāniyya *tamm* and Zenaga *yuktay* 'to continue to', Ḥassāniyya *ʕgab* and Zenaga *yaggara* 'to end up doing'. One of the most notable parallel innovations however, concerns the future morpheme: Ḥassāniyya *lāhi* (invariable participle of an otherwised obsolete verb, but compare *ltha* 'to pass one's time') and Zenaga *yanhāya* (a conjugated verb also meaning 'to busy oneself with something', in addition to its future function). In both cases we have forms related to Classical Arabic *lahā* 'to amuse oneself', with the Zenaga form apparently being a borrowing. It seems, therefore, that this borrowing preceded the *lāhi* of Ḥassāniyya and likely then influenced its adoption as a future tense marker. Note also that in the Arabic dialect of the Jews of Algiers, *lāti* is a durative present tense marker (see **Cohen1924**: 221; Taine-**Cheikh2004**: 224; 2010a: 126–127; 2009: 99).

Ḥassāniyya and Zenaga also display common features with regard to complex phrases. For example, concerning completives, Zenaga differs from other Berber languages in its highly developed usage of *ad/að*, and in particular in the grammaticalized usage of this demonstrative as a quotative particle after verbs of speaking and thinking (Taine-**Cheikh2010a**). This may have had an influence on the usage of the conjunctions *an(n)-* and *ʕan-* (the two forms tend to be confused) in the same function in Ḥassāniyya.

Finally, regarding the variable appearance of a resumptive pronoun in Ḥassāniyya object relative clauses, if influence from Berber (where a resumptive pronoun is always absent) has played any role here, it has simply been to reinforce a construction already attested in the earliest Arabic, whereby the resumptive pronoun is absent if the antecedent is definite, as in (1).

(1)  nṛədd        ʕlī-kum əṛ-ṛwāye lli  ṛadd-Ø        ʕlī-ya muḥammad
     tell.IMPF.1SG on-2PL   DEF-story REL tell.PRF.3MSG-Ø on-1SG Mohammed

     'I am going to tell you the story that Mohammed told me.'

### 3.3.2  Regional influence of Maghrebi Arabic

The Ḥassāniyya spoken in the south of Morocco is rather heavily influenced by other Arabic varieties spoken in the region. Even among those who conserve virtually all the characteristic features of Ḥassāniyya (preservation of interdentals, synthetic genitive expression, absence of the pre-verbal particle *kā-* or *tā-*, absence of discontinuous negation, absence of the indefinite article), particular features of the Moroccan Arabic koiné appear either occasionally or regularly among certain speakers. The most common such features are perhaps the genitive particle *dyal* (Taine-**Cheikh1997b**: 98) and the preverbal particle *kā* (**Aguadé1998**: 211, §**??**; 213, §**??**).

In the Ḥassāniyya of Mali, usage of a genitive particle remains marginal, although **Heath2004** highlights a few uses of genitive *(n)tāʕ* in his texts.

### 3.4  Lexicon

### 3.4.1  Confirmed loanwords

#### 3.4.1.1  Loanwords from standard Arabic

Verbs loaned from standard Arabic are as common as nominal and adjectival loans. Whatever their category, loans are often distinctive in some way (whether because of their syllabic structure, the presence of particular phonemes or their

morphological template), since the lexeme usually (though not always) has the same form in both the recipient language and the source language. Examples of loans without any distinctive features are *barrar̩* 'to justify', and *ðahbi* 'golden'.

A certain number of standard Arabic verbs with the infix *-t-* or the prefix *sta-* are borrowed, but these verbal patterns can be found elsewhere in Ḥassāniyya.

Certain lexical fields exhibit a particularly high degree of loans from standard Arabic: anything connected with Islamic Studies or abstract concepts (religion, rights, morality, feelings etc.) and, more recently, politics, media and modern material culture. These regularly retain the meaning (or one of the meanings) of the source-language item.

### 3.4.1.2 Loanwords from Berber

There many lexical items that are probable loans from Berber, with a number of certain cases among them.

Here we may point to several non-Arabic-origin verbs with cognates across a wide range of Berber languages, such as *kr̩at̩* 'to scrape off' (Zenaga *yugr̩aḏ*); *šayḏaḏ* 'to make a lactating camel adopt an orphaned calf from another mother' (Zenaga *yaṣṣuḏaḏ* 'to breastfeed', *yuḍḍaḏ* 'to suckle'); *santa* 'to begin' (Zenaga *yassanta* 'to begin', Tuareg *ent* 'to be started, to begin'); *gaymar* 'to hunt from a distance' (Berber *gmǝr* 'to hunt').

Other verbs are derived from nouns loaned from Berber. Hence, *γawba* 'to restrain a camel, put it in an *aγāba*' (Tuareg *aγaba* 'jaws'). Sometimes there is both a verb and an adjective stemming from a loaned root, as in *gaylal* 'to have the tail cut' and *agīlāl* 'having a cut tail' (Tuareg *gilel* and *agilal*).

Some loaned Ḥassāniyya nouns are found with the same root in Berber languages other than Zenaga. For example, *agayš* 'male bustard' (Tuareg *gayǝs*), *āškǝr̩* 'partridge' (Kabyle *tasekkurt* in the feminine form), *tayffārǝt* 'fetlock (camel)' (Zenaga *tiʔffart*, Tuareg *téffart*), *(n)tūrža* '*Calotropis procera*' (Zenaga *turžah,* Tuareg *tǝrza*), *talawmāyǝt* 'dew' (Zenaga *tayaṃut*, Tuareg *tălǎmut*); *azāγǝr* 'wooden mat ceiling between beams' (Zenaga *azaγri* 'lintel, beam (of a well)', Tuareg *ǝzgǝr* 'to cross', *ǎzǝgǝr* 'crossbeam').

Most of the loanwords cited above are attested in Zenaga (sometimes in a more innovative form than is found in other Berber varieties, such as *yaggīyyay* 'to have a cut tail' where /y/ < *l). However, there are numerous cases where a corresponding Berber item is attested only in Zenaga. In such cases it is difficult to precisely identify the source language, even if the phonology and/or morphology seems to indicate a non-Arabic origin.

Loanwords from Berber seem to be particularly common in the lexicon of fauna, flora, and diseases, as well as in the field of traditional material culture (objects, culinary traditions, farming practices etc., see Taine-**Cheikh2010b**; 2014). Unlike the form of the loans, which is often quite divergent from that of the source items, their semantics tends to remain largely unchanged. However, there are some exceptions, notably when the verbs have a general meaning in Berber (see "to breastfeed").

### 3.4.1.3 Loanwords from Sahel languages

Some Ḥassāniyya lexical items seem to be borrowed directly from African languages, though the precise origin is only known in a few cases. We may note, in addition to *gad^y* 'dried fish' (< Wolof) and *d^yəngra* 'warehouse' (< Soninke), a few terms which appear to be borrowed from Pulaar: *ḅamḅa* 'to carry a child on one's back', *t^yəhli* 'roof on pillars' and *kīri* 'boundary between two fields'.

In some regions we find a concentration of loans in particular domains in relation to specific contact languages. For example, in the ancient town of Tichitt, we find borrowings from Azer and Soninke (Jacques-**Meunié1961**; **Monteil1939**; **Diagana2013**): *kā* 'house' (Azer *ka/kany*, Soninke *ká*) in *kā n laqqe* 'entrance of the house'; *killen* 'path' (Azer *kille*, Soninke *kìllé*); *kunyu/kenyen* 'cooking' (Azer *knu/kenyu*, Soninke *kìnŋú*).

A significant list of loanwords from Songhay has been compiled by **Heath2004** in Mali, including e.g.: *ṣawṣab* (< *sosom/sosob*) 'pound (millet) in mortar to remove bran from grains'; *daydi/dayday* (< *deydey*) 'daily grocery purchase'; *ākāṛāy* (< *karey*) 'crocodile'; *sari* (< *seri*) 'millet porridge'. Only *sari* has been recorded elsewhere in Mauritania (in the eastern town of Walata). On the other hand, all authors who have done field work on the Ḥassāniyya of Mali (particularly in the region of Timbuktu and the Azawad), have noted loanwords from Songhay. This is true also of **Clauzel1960** who, as well as a number of Berber loanwords, gives a small list of Songhay-derived items used in the salt mine of Tāwdenni, such as *titi* 'cylinder of saliferous clay used as a seat by the miners' (< *tita*) and *t^yar* 'adze' (< *t^yara*).

### 3.4.1.4 Loanwords from Indo-European languages

The use of loanwords from European languages tends to vary over time. Thus, a large proportion of the French loanwords borrowed during the colonial period have more recently gone out of use, such as *bəṛṭmāla* or *qoṛṭmāl* 'wallet' (< *porte-monnaie*), *dabbīš* 'telegram' (< *dépêche* 'dispatch') or *ṣaṛwaṣ* 'to be very close to

the colonizers' (< *service* 'service'). This is true not only of items referring to obsolete concepts (such as the currency terms *sūvāya* 'sou' or *ftən/vəvtən* 'cent', likely < *fifteen*), but also of those referring to still-current concepts which are, however, now referred to with a term drawn from standard Arabic (e.g. *ministr* 'minister', replaced by *wazīr*). This does not, however, eliminate the permanence of some old loanwords such as *wata* 'car' (< *voiture*) or *maṛṣa* 'market' (< *marché*).

Although not unique to Ḥassāniyya, the frequency of the emphatic phonemes (especially /ṣ/ and /ṭ/) in loans from European languages is notable. Consider, in addition to the treatment of *service*, *porte-monnaie* and *marché* as noted above, that of *baṭṛūn* 'boss' (< *patron*), which gives rise to *tbaṭṛan* 'to be(come) a boss' *ṭawn* 'ton' (< *tonne*).

Ould Mohamed **Baba2003** gives an important list of loanwords from French and offers a classification by semantic field.

### 3.4.2 More complex cases

#### 3.4.2.1 *Wanderwörter*

Various Arabic lexical items are of Latin, Armenian, Turkish, Persian, origin and so on. Whether we are dealing with the names of calendar months, or of items such as trousers (*sərwāl)*, these terms are not borrowed directly from the source language by Ḥassāniyya and are found elsewhere (e.g. *balbūẓa* 'eyeball' < Latin *bulbus*, attested throughout the Maghreb). The history of such items will not be dealt with here. We can, however, mention the case of some well-attested terms in Ḥassāniyya that appear to have been borrowed from sub-Saharan Africa.

One such is *māṛu* 'rice', which seems to come from Soninke (*máarò*), although it is also attested in Wolof (*mālo*) and Zenaga (*mārih*). Another term, which is just as emblematic, is *mbūṛu* 'bread', whose origin has variously been attributed to Wolof, Azer, Mandigo (and even English *bread*).

To these very everyday terms, we may also add *ṃutri* 'pearl millet' and *makka* 'maize', which have the same form both in Ḥassāniyya and in Zenaga. The first is a loanword from Pulaar (*muutiri*). The second is attested in many languages and seems to have come from the placename Mecca.

As for *garta* 'peanut', *ḷāḷo/ḷaḷu* 'pounded baobab leaves that serve as a condiment' (synonym of *taqya* in the southwest of Mauritania) and *kəddu* 'spoon', these appear to be used just as frequently in Pulaar as they are in Wolof.

### 3.4.2.2 Berberized items

Despite the absence of any Berber affixes in the list of loanwords in, only *kəddu* 'spoon' is regularly used with the definite article. In this regard, these loanwords act like words borrowed from Berber, or more generally, those with Berber affixes.

It is, in fact, difficult to prove that a noun with this kind of affix is definitely of Berber origin, since we find nouns of various origins with Berber affixes. Some of them are loanwords from the languages of the sedentary people of the valley, such as *adabāy* 'village of former sedentary slaves (*ḥṛāṭīn*)' (< Soninke *dèbé* 'village'); *iggīw/īggīw* 'griot' (Zenaga *iggiwi*, borrowed from Wolof *gēwel* or from Pulaar *gawlo*). Others are borrowed from French: *agāṛāž* 'garage'; *təmbīskit* 'biscuit'. Even terms of Arabic origin are Berberized, as is likely the case with with *tasūvra* 'large decorated leather bag for travelling' (cf. *sāvər* 'to travel') or *tāẓəẓmît* 'asthma' (cf. CA *zažma* 'shortness of breath when giving birth').

### 3.4.2.3 Reborrowings

Instances of back and forth between two languages – primarily Ḥassāniyya and Zenaga – seem to be the reason for another type of mixed form, illustrated previously in 3.2.2.1. by the Zenaga verbs *yassəðbah* 'to leave in the afternoon' and *yišnar* 'to orient oneself'.

Ḥassāniyya *saγnan* 'to mix gum with water to make ink' provides another example where this time the points of departure and arrival seem to be from the Arabic side. In fact, this loanword is a borrowing of Zenaga *yassuγnan* 'to thicken (ink) by adding gum', a verb formed from *əssaγan* 'gum'. This noun in turn appears to be an adaptation of the Arabic *samγa* 'ink'.

In the case of *sla* 'placenta': there is a double round-trip between the two languages, this time without metathesis: after a passage from Arabic to Zenaga (> *əs(s)la*), there is return to Ḥassāniyya with the causative verb *sasla* 'to soak a hide' and a second borrowing into Zenaga with the reflexive form *(yə)stasla* 'start to lose fur (of soaked hides)'.

### 3.4.2.4 Calques

Calques are undoubtedly common, but they are particularly frequent in locutions such as *rəggət əž-žəll* 'susceptibility' and *bū-damʕa* 'rinderpest' (literally 'thinness of skin' and 'the one with a tear'). *taššəddi-n əyim* and *ən-anḍi* are, in fact, exact calques in Zenaga (Taine-**Cheikh2008a**).

### 3.4.2.5 Individual variation

The permeability of loanwords differs from one individual to another. This is natural when we are dealing with bilingual speakers and this probably explains the special features of the Ḥassāniyya of the Awlād Bann^yūg (often bilingual speakers of Ḥassāniyya and Wolof) or the Ḥassāniyya of Mali (where Arabic speakers often speak Songhay and sometimes Tamasheq). However, it also depends on the individuals in question in terms of their "loyalty" to the language, whether the language is under the pressure of Moroccan Arabic koiné in Morocco (Taine-**Cheikh1997b**; **Heath2002**; **Paciotti2017**), or whether it is imposed as a lingua franca in Mauritania (**Dia2007**).

## 4  Conclusion

The principal domain affected by the contact in Ḥassāniyya is that of the lexicon (though an assessment in percentage terms is not at present possible). However, the integration of loanwords – in particular those from standard Arabic and Berber – has resulted in a significant enrichment of the phonological system and of the inventory of nominal patterns. The effects of contact on the verbal morphology and syntax of the dialect are more indirect. The major developments in Ḥassāniyya seem most likely to instead be a product of internal evolution. In certain cases, Zenaga has probably had an influence; in others, we rather witness instances of parallel evolution.

In future, by studying the vehicular Ḥassāniyya of Mauritania and of the border regions (southern Morocco, southern Algeria, Senegal, Niger, and so on) we will perhaps discover new developments as a result of contacts triggered by the societal and political changes of the 21st century.

## Further reading

Links between Ḥassāniyya and other languages are particularly complex at the level of semantics and lexicon. On these topics, beyond the available Ḥassāniyya and Zenaga dictionaries (**Heath2004**; Taine-**Cheikh1988**–1998; 2008b), readers may consult the available studies of specific fields (**Monteil1952**; Taine-**Cheikh2013**) or particular template (Taine-**Cheikh2018b**).

## Abbreviations

Berb.= Berber; CA=Classical/standard Arabic; Hass.=Ḥassāniyya; s.o. =someone;
Zen.=Zenaga

```
%%move bib entries to localbibliography.bib
\end{stylelsUnNumberedSection}
```

```
@article{Aguadé1998,
author = {Aguadé, Jordi},
journal = {\textit{Estudios} \textit{de} \textit{Dialectología} \textit{Norteafr
pages = {203–215},
sortname = {Aguade, Jordi},
title = {Relatos en hassaniyya recogidos en Mħāmīd (Valle del Dra, Sur de Marrue
volume = {3},
year = {1998}
}
```

```
@book{Cantineau1960,
address = {\textit{Études} \textit{de} \textit{linguistique} \textit{arabe:} \te
125. Paris},
author = {Cantineau, Jean},
publisher = {Klincksieck},
title = {Cours de phonétique arabe},
year = {1960}
}
```

```
@book{Clauzel1960,
address = {Institut de recherches sahariennes de l'Université d'Alger},
author = {Clauzel, Jean},
publisher = {Imprimerie Protat frères},
title = {\textit{L'exploitation} \textit{des} \textit{salines} \textit{de} \text
```

```
year = {1960}
}




@book{Cohen1963,
address = {Paris},
author = {Cohen, David},
publisher = {Klincksieck},
title = {\textit{{Le}} \textit{dialecte} \textit{arabe} ḥassānīya \textit{de} \tex
year = {1963}
}




@book{Cohen1924,
address = {Paris},
author = {Cohen, Marcel},
publisher = {Ernest Leroux},
title = {\textit{{Le}} \textit{système} \textit{verbal} \textit{sémitique} \textit
year = {1924}
}




@incollection{Dia2007,
address = {London},
author = {Dia, Alassane},
booktitle = {\textit{{Arabic}} \textit{in} \textit{the} \textit{city:} \textit{Iss
editor = {Catherine Miller, Enam Al-Wer, Dominique Caubet and Janet C. E. Watson},
pages = {325--344},
publisher = {Routledge},
title = {Uses and attitudes towards {Hassaniyya} among Nouakchott's Negro-Mauritan
year = {2007}
}
```

```
@book{Diagana2013,
address = {Paris},
author = {Diagana, Ousmane},
publisher = {Karthala},
title = {\textit{Dictionnaire} \textit{soninké-français} \textit{(Mauritanie)}},
year = {2013}
}
```

```
@book{Heath2002,
address = {London and New York},
author = {Heath, Jeffrey},
publisher = {Routledge},
title = {\textit{{Jewish}} \textit{and} \textit{Muslim} \textit{dialects} \texti
year = {2002}
}
```

```
@book{Heath2004,
address = {Wiesbaden},
author = {Heath, Jeffrey},
publisher = {Harrassowitz Verlag},
title = {\textit{Hassaniya} \textit{{Arabic}} \textit{({Mali})–{English}–
{French}} \textit{Dictionary}},
year = {2004}
}
```

@book{Jacques-Meunié1961,
address = {Paris},
author = {Jacques-Meunié, Djinn},
publisher = {Klincksieck},
sortname = {Jacques-Meunie, Djinn},
title = {\textit{Cités} \textit{anciennes} \textit{de} \textit{Mauritanie}},
year = {1961}
}


@incollection{Monteil1939,
address = {Paris},
author = {Monteil, Charles},
booktitle = {\textit{Contributions} \textit{à} \textit{l'étude} \textit{du} \texti
editor = {Théodore Monod},
pages = {213–341},
publisher = {Larose},
title = {{La} langue azer},
year = {1939}
}


@book{Monteil1952,
address = {Saint-Louis (Senegal)},
author = {Monteil, Vincent-Mansour},
publisher = {Centre IFAN–Mauritanie},
title = {\textit{Essai} \textit{sur} \textit{le} \textit{chameau} \textit{au} \text
year = {1952}
}


@incollection{OuldMohamedBaba2003,

*??*

address = {Cadiz},
author = {Ould Mohamed Baba, Ahmed-Salem},
booktitle = {\textit{{AIDA}} \textit{5th} \textit{Conference} \textit{Proceeding
editor = {Ignacio Ferrando and Juan José Sanchez Sandoval},
pages = {61–74},
publisher = {Universidad de Cádiz, Servicio de Publicaciones},
title = {Emprunts du dialecte ḥassāniyya à la langue française},
year = {2003}
}


@book{Paciotti2017,
address = {La situazione della Ḥassāniyya nella regione di Guelmime-Oued Noun. N
author = {Paciotti, Luca},
publisher = {Università degli Studi di Napoli "L'Orientale" Masters},
title = {Conflitti linguistici e glottopolitica in Marocco},
year = {2017}
}


@book{Taine-Cheikh1988,
address = {Paris},
author = {Taine-Cheikh, Catherine},
publisher = {Geuthner},
title = {\textit{Dictionnaire} \textit{ḥassāniyya–français}},
year = {1988}
}


@article{Taine-Cheikh1997,
author = {Taine-Cheikh, Catherine},
journal = {\textit{Matériaux} \textit{arabes} \textit{et} \textit{sudarabiques}

```
pages = {93–142},
title = {Les emprunts au berbère zénaga: {{{U}}}n sous-système vocalique du ḥassān
volume = {8},
year = {1997a}
}




@article{Taine-Cheikh1997,
author = {Taine-Cheikh, Catherine},
journal = {\textit{Peuples} \textit{méditerranéens}},
pages = {85–102},
title = {Les hassanophones du Maroc: {{{E}}}ntre affirmation de soi et auto-renieme
volume = {79},
year = {1997b}
}




@incollection{Taine-Cheikh2003,
address = {Cadiz},
author = {Taine-Cheikh, Catherine},
booktitle = {\textit{{AIDA}} \textit{5th} \textit{Conference} \textit{Proceedings}
editor = {Ignacio Ferrando and Juan José Sanchez Sandoval},
pages = {103–118},
publisher = {Universidad de Cádiz, Servicio de Publicaciones},
title = {Les valeurs du préfixe \textit{s-} en hassaniyya et les conditions de sa g
year = {2003}
}




@article{Taine-Cheikh2004,
author = {Taine-Cheikh, Catherine},
journal = {Réflexions pour une typologie. \textit{Estudios} \textit{de} \textit{Di
```

```
pages = {215–238},
title = {{Le}(s) futur(s) en arabe},
volume = {8},
year = {2004}
}




@book{Taine-Cheikh2008,
address = {\textit{Berber} \textit{in} \textit{contact:} \textit{Linguistic} \te
138. Cologne},
author = {Taine-Cheikh, Catherine},
publisher = {Rüdiger Köppe},
title = {Arabe(s) et berbère en contact: {{L}}e cas mauritanien},
year = {2008a}
}




@book{Taine-Cheikh2008,
address = {Cologne},
author = {Taine-Cheikh, Catherine},
publisher = {Rüdiger Köppe},
title = {\textit{Dictionnaire} \textit{zénaga–français.} \textit{{Le}} \textit{b
year = {2008b}
}




@article{Taine-Cheikh2009,
author = {Taine-Cheikh, Catherine},
journal = {\textit{Faits} \textit{de} \textit{Langue}},
pages = {91–102},
title = {Les morphèmes de futur en arabe et en berbère: {{R}}éflexions pour une
volume = {33},
```

```
year = {2009}
}




@incollection{Taine-Cheikh2010,
address = {Amsterdam},
author = {Taine-Cheikh, Catherine},
booktitle = {\textit{Clause} \textit{linking} \textit{and} \textit{clause} \textit
editor = {Isabelle Bril},
pages = {355–398},
publisher = {John Benjamins},
title = {The role of the Berber deictic and {TAM} markers in dependent clauses in
year = {2010a}
}




@misc{Taine-Cheikh2010,
author = {Taine-Cheikh, Catherine},
note = {Réflexions à partir des lexiques arabes et berbères. \textit{The} \textit{
2). 64–88.},
title = {Aux origines de la culture matérielle des nomades de Mauritanie},
year = {2010b}
}




@incollection{Taine-Cheikh2013,
address = {Louvain},
author = {Taine-Cheikh, Catherine},
booktitle = {\textit{In} \textit{and} \textit{out} \textit{of} \textit{{Africa}:} \
editor = {Carole de Féral},
pages = {137–164},
publisher = {Peeters},
```

title = {Des ethnies chimériques aux langues fantômes: {{L}}'exemple des Imrague
year = {2013}
}

@article{Taine-Cheikh2014,
author = {Taine-Cheikh, Catherine},
journal = {\textit{Awal—Cahiers} \textit{d'études} \textit{berbères}},
pages = {27–50},
title = {Les voies lactées: {{L}}e lait dans l'alimentation des nomades de Mauri
volume = {42},
year = {2014}
}

@book{Taine-Cheikh2016,
address = {Rabat},
author = {Taine-Cheikh, Catherine},
publisher = {Centre des Études Sahariennes},
title = {\textit{Études} \textit{de} \textit{linguistique} \textit{ouest-saharie
year = {2016}
}

Taine-Cheikh, Catherine. 2018a. Historical and typological approaches to Maurita

@incollection{Taine-Cheikh2018,
address = {Amsterdam},
author = {Taine-Cheikh, Catherine},
booktitle = {\textit{{Arabic}} \textit{in} \textit{contact}},

```
editor = {Stefano Manfredi and Mauro Tosco},
pages = {135–159},
publisher = {John Benjamins},
title = {Ḥassāniyya {Arabic} in contact with Berber: {{T}}he case of quadriliteral
year = {2018b}
}
```

# Chapter 14

# Contact and variation in Arabic intonation

??

??

Evidence is emerging of differences among Arabic dialects in their intonation patterns, along known parameters of variation in prosodic typology. Through a series of brief case studies, this chapter explores the hypothesis that variation in intonation in Arabic results from changes in the phonology of individual Arabic varieties, triggered by past (or present day) speaker bilingualism. If correct, variation in intonation should reflect prosodic properties of the specific languages that a particular regional dialect has had contact with.

## 1 Introduction

### 1.1 Contact-induced variation in intonation

The hypothesis explored in this chapter is that observed synchronic variation in intonation across Arabic dialects is contact-induced. In this scenario, differences between dialects would result from changes in the intonational phonology of individual varieties triggered by speaker bilingualism in Arabic and one or more other languages (**Lucas2014**), either in the past, or up to and including the present day. To achieve this, we outline a framework for analysis of variation in intonation (§??), summarise recent research on the effects of bilingualism on the intonational phonology of bilingual individuals and the languages they speak (§??) and sketch the types of language contact scenario which may be relevant for Arabic (§??). In §?? we present case studies of prosodic features which appear to be specific to a particular dialect, on current evidence at least, and discuss which of the potentially relevant contact languages might have served as the potential

source of the feature in question, considering also possible endogenous (internal) sources of the change. The chapter closes (§**??**) with suggestions for future research.

## 1.2 Cross-linguistic variation in intonation

Any attempt to delimit the nature and scope of variation in intonation depends on the model of intonational phonology adopted. The analyses explored in §**??** below are framed in the Autosegmental-Metrical (AM) theory of intonation (**Ladd2008**), and the parameters of intonational variation explored are thus influenced by this choice.

A basic debate in analysis of intonation is whether the primitives of the system are whole contours (defined over an intonational phrase), or some sub-component of those contours (**Ladd2008**). In AM theory, intonation is modelled as interpolation of pitch between tonal targets; these tonal targets are the primitives of the system and are of two types: pitch accents are associated with the heads of metrical domains (e.g. stressed syllables), boundary tones are associated with the edges of metrical domains (e.g. prosodic phrases). In AM tonal targets are transcribed using combinations of high (H) or low (L) targets, which equate to significant peaks and valleys, respectively, in the pitch contour of the utterance; association of these events to landmarks in the metrical structure is marked using '*' for pitch accents (associated with stressed syllables) and '%' for boundary tones (associated with the right edge of prosodic phrases of different sizes). A typical AM analysis yields an inventory of the pitch accents and boundary tones needed to model the contours in a corpus of speech data, supported by a description of the observed contours (**JunFletcher2014**).

**Ladd2008**'s (**Ladd2008**) taxonomy of possible parameters of cross-linguistic variation in intonation (based on **Wells1982**) envisages four broad (inter-related) categories of variation: systemic (differences in the inventory of pitch accents or boundary tones), semantic (differences in the meaning or function associated with a particular contour, pitch accent or boundary tone), realisational (differences in the phonetic realisation of otherwise parallel pitch accents or boundary tones) and phonotactic (differences in the distribution of pitch accents and boundary tones, or in their association to metrical structure).

Comparison of AM analyses across a typologically distinct set of languages (**Jun2005**; 2014) has highlighted systematic cross-linguistic variation of a systemic and/or phonotactic nature, in terms of prosodic phrasing (with relatively smaller or larger domains involved in structural organization of intonation patterns), the distribution of tonal events relative to prosodic constituents (mark-

ing either the edges or the metrical heads of phrases or both), and the size and composition of the inventory of tonal events regularly observed (pitch accents and boundary tones). There is also a large body of research on cross-linguistic variation in the phonetic realisation of pitch accents, in particular on peak alignment (**AttererLadd2004**; **Ladd2006**) and scaling (**LaddMorton1997**), confirming the existence of realisational cross-linguistic variation. The most advanced work on semantic variation to date has been on Romance languages, facilitated by a concerted effort to develop descriptions of their intonation patterns within a common annotation system (**FrotaPrieto2015**).

For Arabic, evidence is emerging of variation along similar lines. Recent review articles have highlighted clear differences in the size and composition of the inventory of pitch accents and boundary tones across Arabic dialects (**Chahal2006**; El **Zarka2017**), and in the association of pragmatic meanings with contours (cf. the case study in §**??**). Initial evidence suggests a difference between Jordanian and Egyptian Arabic in the mapping of prosodic phrases to syntax (**Hellmuth2016**) similar to that reported across Romance languages (D'**ImperioEtAl2005**). Recent research suggests that Moroccan Arabic is a non-head-marking language in contrast to other Arabic dialects which are head-marking (see §**??**), mirroring the cross-linguistic variation captured in **Jun2005**'s (**Jun2005**) typology, and among the head marking dialects, there appears to be variation in the density of distribution of pitch accents (**ChahalHellmuth2015**) (see §**??**).

## 1.3 Contact-induced variation in intonation

A growing body of research has explored contact-induced prosodic change in the speech of bilingual communities and individuals. The initial focus of most studies was on second language (L2) learners' intonation patterns, or studies of individual bilinguals (**Queen2012**), and early L2 studies focused on realisational effects of a speaker's L1 on their L2, and vice versa (**AttererLadd2004**; **Mennen2004**). More recent studies reveal a complex array of prosodic effects, both in terms of the features involved in the change (taking in all four of Ladd's categories of possible variation), and also in the directionality of effects (L1 on L2, L2 on L1, or hybrid effects).

**Bullock2009** characterizes the general contact-induced language change literature (**WeinreichEtAl1968**; **ThomasonKaufman1992**) as having made the assumption that segmental effects would 'precede' prosodic effects, thus predicting prosodic effects would be seen only in contexts of widespread or sustained community bilingualism. As Bullock notes, however, there is no logical structural reason why this should be the case; her own study of English-like prosodic

patterns in heritage French speakers in Pennsylvania confirms an effect of the dominant language in the prosodic domain (specifically in the realisation of focus) in speakers who in other respects maintain French segmental patterns.

Another example of prosodic properties of a dominant language affecting prosodic realisation of a heritage or second language, is that of immersion Gaelic learners in Scotland (**Nance2015**). Nance demonstrates a structural change in progress in Gaelic, from lexical pitch accent – still used by older English-Gaelic bilinguals – to a purely post-lexical system, used by younger bilinguals in immersion education who produce Gaelic with English-like intonation. Similar effects of the dominant language on the non-dominant language are reported for Spanish in contact with Quechua (O'**Rourke2004**).

The reverse effect has also been found in a number of studies, however, where prosodic properties of a non-dominant or heritage language have an effect on the prosodic realisation of the dominant language, in the speech of an individual or of the whole community. **Fagyal2005** studied a group of bilingual French-Arabic adolescents in Paris; instead of a typical French phrase-final rise, these speakers produce a phrase-final rise-fall contour in declaratives similar to the contour observed in Moroccan Arabic (MA) in parallel contexts. **Simonet2011** shows that the steep ('concave') final fall in Majorcan Catalan declaratives is now widely observed in Majorcan Spanish, replacing the typical gradual ('convex') fall in Majorcan Spanish, but that each individual bilingual's usage closely mirrors their reported language dominance. **ColantoniGurlekian2004** observe patterns of peak alignment in pre-nuclear accents and pre-focal downstep in Buenos Aires Spanish which differ from neighbouring varieties of Spanish but resemble those in Italian, and ascribe them to high levels of Spanish-Italian bilingualism in the city in the late 19th/early 20th century. In this last case, the period of community bilingualism which triggered the change is now long past, but the effect on the prosodic patterns of the dominant language (in this case, Spanish) persists.

Finally, Queen (2001; 2012) reports a case of 'fusion': Turkish-German bilinguals in Germany display phrase-final intonation contours which are never used by monolinguals in either language, but are found only in the speech of bilinguals, in a new variety of German known as 'Türkisch-Deutsch'.

This emerging literature suggests that contact-induced prosodic change is a frequent phenomenon, arising in varied forms and across diverse contact situations. **Bullock2009** suggests that prosody and intonation are especially prone to change for three reasons. First, because the acoustic parameters involved - pitch, intensity and duration - are part of the linguistic encoding of all languages, albeit in different constellations, and are thus readily adapted. Second, because, percep-

tually, all languages make use of prosodic parameters to convey some aspects of utterance-level meaning, thus the mapping of form to meaning is also readily adapted. Third, and perhaps most persuasively, because the form-meaning mapping in intonation is generally not fixed, but displays considerable inter- and intra-speaker variation (**CangemiEtAl2015**; **CangemiEtAl2016**) as well as contextual variation (cf. **Walker2014**), and it is pockets of structural 'indeterminacy' of this kind which are prone to change in bilingual grammars (**Sorace2004**). **Queen2001** also suggests that the intertwining of form and function in intonation makes it a fruitful sphere for investigation of contact-induced change, because "intonation is one of the few linguistic elements that comments simultaneously on grammar, context and culture". Indeed, **Simonet2011**'s (**Simonet2011**) work shows that speakers are able to adopt the intonation of a contact language without actually being proficient in the source language. Finally, **Matras2007** argues that the separate nature of prosody, which is processed separately from segmental phonology, and can be interpreted independently of the propositional content of the utterance, renders prosody more 'borrowable' than other aspects of the grammar.

In sum, there is strong evidence that intonation patterns are highly porous, being transferred between dominant and non-dominant languages in either direction; intonation is thus a fruitful area for investigation of contact-induced language change. Among the literature reviewed here, the paper by **ColantoniGurlekian2004** most closely resembles the type of work which will be needed in future for Arabic; they investigate present day intonational variation in closely related varieties, and provide evidence from historical migration patterns to support the claim that the present day variation can be ascribed to an earlier period of widespread bilingualism, in a language which is a plausible source of the feature in question. In the next section we outline a similar line of investigation for Arabic.

## 1.4 Contact-induced variation in Arabic

The time depth of descriptions of intonation patterns is shallow, due to the lack of historical audio recordings, and a general tendency that traditional grammars do not include detailed descriptions of prosody. It is thus difficult to reliably determine when changes in intonation may have happened, and the range of languages to be considered as the source of any putative intonational change is rather broad.

One set of potential source languages is the substrate languages spoken in a particular region before the arrival of Arabic, for example Amazigh (Berber) in North Africa, and Coptic and Egyptian in Egypt. We might also see influence

from the external languages which these indigenous languages were in contact with prior to the arrival of Arabic, such as Greek and Latin, or of other external languages whose influence was felt throughout the Arab world in later periods, such as Persian and Ottoman. Other possible source languages are European languages spoken along the northern coast of the Mediterranean, since large areas of southern Europe were under Arab rule for extended periods (7th-15th centuries CE), and contact through sea-borne trade is likely to have continued after that time. Conversely, large areas of the Arab world were under direct or indirect European control also (19th -20th centuries CE) and the influence of these languages is still felt today. Finally, we might also consider the potential effects of contact with global languages such as English, and with the L1 languages of migrant workers and long-term displaced language communities.
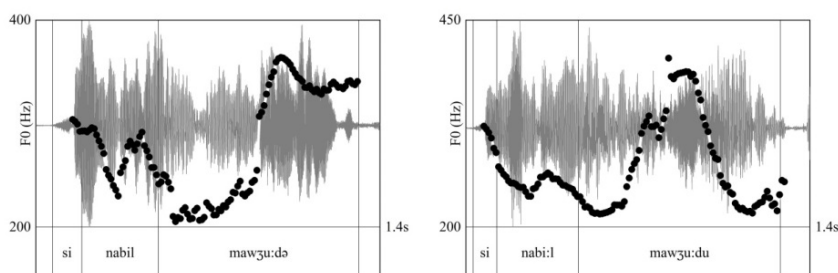
The decision to treat observed present day variation as the result of change does not entail assuming that any one variety of Arabic was the ancestor of all dialects. Instead, the approach here will be to identify prosodic features which are seen in one Arabic dialect (or group of dialects), but not (yet) seen in any other dialects, as the most likely cases of potential contact-induced change. In each case study we evaluate the hypothesis by looking for evidence of the same feature in the relevant contact languages for the dialect in question, with comparison to possible language internal sources of the change.

## 2 Contact-induced variation in Arabic intonation

### 2.1 Tunisian Arabic question marking

Tunisian Arabic polar questions are typically associated with a salient rise-fall pitch contour at the end of the utterance: speakers from south-east Tunisia produce a complete rise-fall in which pitch rises over the stressed syllable of the last word in the utterance to a peak, then falls to low; in contrast, speakers from Tunis produce a 'rise-plateau' contour, in which, after the peak, pitch falls slightly then levels out. These patterns are illustrated in Figure 1 (Bouchhioua et al. in press). The rise-fall prosodic contour is frequently accompanied by a segmental question marker, in the form of a vowel added to the end of the last word in the utterance. The quality of the epenthesised vowel is influenced partly by vowels earlier in the word (in a form of vowel harmony) and partly by regional dialect, though these vowel quality patterns require further investigation.

The rise-fall yes/no-question contour in TA differs from the rise seen in yes/no-questions in most Arabic dialects (Hellmuth to appear) and, in terms of distri-

please provide either eps or png of this graphic. In its present form, the graphic cannot be printed

| si | naˈbil | /mawˈʒud/ [mawˈʒudə/u] |
|----|--------|------------------------|
| Mr | Nabil | present |

'Is Mr Nabil there?' (tuno-arc1-f1/tuse-arc1-f1)

Figure 1: : Pitch trace of yes-no questions from a Tunisian Arabic speaker from Tunis (left) and from the south-east (right).

bution, from the rise-fall contour observed in Moroccan Arabic (MA) across all utterance types (not only in yes/no-questions). The vowel epenthesis marker appears to be unique to TA, thus far.

A pattern of utterance-final vowel epenthesis has been observed in a number of Romance languages spoken along the northern edge of the Mediterranean, including Bari Italian (**GriceEtAl2015b**), and different varieties of Portuguese (**FrotaEtAl2015**). These cases of utterance-final vowel epenthesis are interpreted as 'text-tune adjustment', where segmental material is added to accommodate a complex prosodic contour. For example, in Standard Portuguese, a more general rule of utterance-final vowel deletion is blocked in utterances bearing a complex prosodic contour, such as the fall-rise (H+L* LH%) on yes/no-questions (**FrotaEtAl2015**). In Bari Italian, epenthesis is seen on a range of utterance types, but – like Portuguese – its occurrence can be ascribed to 'tonal crowding' (i.e. that the complex contour requires more segmental material to be realised). This is reflected in higher incidence of epenthesis on utterance-final monosyllables than on longer words, and on words in which the final sound is an obstruent than on words with a final sonorant (**GriceEtAl2015b**).

Investigation of utterance-final vowel epenthesis in TA yes/no-questions, in

a corpus of data collected in Tunis, shows a very different pattern, however. In TA the incidence of epenthesis is not affected by the number of syllables in the utterance-final word nor by the type of final sound. In addition, whereas in the other Romance languages epenthesis occurs on a range of utterance types, in TA epenthesis occurs only in yes/no-questions, and predominantly in yes/no-questions which are produced with a complex rise-plateau or rise-fall contour (Hellmuth in press-b). The effects which in the Romance languages are taken as evidence of text-tune adjustment are lacking in TA, which appears to rule out a language internal (endogenous) source of the TA pattern of vowel epenthesis.

The TA epenthetic vowel is in fact best characterized as an optional question marker comprising the vowel itself plus an accompanying fall in pitch. The segmental marker is well-known among Tunisian linguists, being described as "the pan-Tunisian question marker clitic [ā]" (**HerinZammit2017**: 141), but the accompanying prosodic contour has received little attention in the literature until recently. This traditional question marking strategy may however be in decline, since it now alternates with realisation of a yes/no-question using a simple rise contour similar to that found in most other Arabic dialects, and without an utterance-final epenthetic vowel. The possibility of a change in progress is suggested by the somewhat higher incidence of the epenthesis among young female speakers (Hellmuth in press-b).

The epenthesis + complex contour strategy in TA yes/no-questions stands out from other Arabic dialects and may thus be due to contact-induced prosodic change. Italian was spoken in Tunisia more widely than French, in the late 19th century (**Sayahi2011**), and is thus a potential source of the contour, since rise-falls occur in yes/no-questions in a number of Italian dialects (Gili **FivelaEtAl2015**). However, the conditioning environments of epenthesis reported for Bari Italian are very different, suggesting that contact with Italian is not a likely source of the epenthesis component of the TA pattern.

An alternative source of the vowel epenthesis pattern is French, since Tunisia has seen very high levels of bilingualism in TA and French from the late 19th century up to the present day, despite concerted efforts to reduce usage of French (**Daoud2007**). Utterance-final schwa epenthesis has been reported as an emerging phenomenon in French (**Hansen1997**), but its distribution is again much broader, being seen across a range of utterance types, and not restricted to yes/no-questions. Despite clear evidence of contact-induced effects of French on TA in other domains, such as lexical borrowing, and a general trend towards use of French by female speakers (**Walters2011**), the different distribution of final epenthesis in French suggests it is not the most likely source of the TA segmental question

marking strategy.

The other major contact language with TA is Tunisian Berber (TB). Although levels of TA-Berber bilingualism in Tunisia are now low, other than in certain regions (**Gabsi2011**), there was a sustained period of TA-TB bilingualism from the 11th century and TB is an important substrate of TA (**Daoud2007**). Although there are no studies of the prosody of TB, to our knowledge, a recent detailed study of Zwara Berber, spoken close to the Tunisian border in western Libya, documents a polar question marking clitic /a/ which is obligatorily accompanied by a rise-fall contour (**Gussenhoven2017**). The match of this description to the TA pattern is so close that it seems plausible that the TA question marking pattern arose due to contact with TB during the period of sustained TA~TB bilingualism. The greater use of the epenthesis + contour strategy by female speakers than male speakers, as well as regional variation, makes this feature of TA ripe for further detailed sociolinguistic study.

## 2.2 Moroccan Arabic word prosody

Variation in word stress patterns across Arabic dialects has inspired much phonological investigation (**Watson2011**), but the Moroccan Arabic (MA) stress system has defied analysis until recently. **Mitchell1993** notes that "in contrast with all the other vernaculars […], the place of prominence in a word in isolation is not carried over to its occurrence in the phrase and sentence", and this characterization was confirmed experimentally by **Boudlal2001**. A range of positions have emerged, with some authors claiming that MA does have word stress (**Benkirane1998**; **BurdinEtAl2014**), and others that it does not (**Maas2013**; El **Zarka2012**).

It is now clear that MA is indeed typologically different from most other Arabic dialects in its word prosody. Whereas the majority of Arabic dialects have salient word-level stress and are thus clearly 'head-marking' languages, in the typology proposed by **Jun2005**, MA is a non-head-marking language in which tonal events mark the edges of prosodic phrases only. **Bruggemann2018** provides acoustic evidence that there are no consistent cues to lexical prominence in MA, and perceptual evidence that MA listeners display the same type of 'deafness' to stress as has been reported for listeners in languages which also lack head-marking such as French (**DupouxEtAl2001**) and Persian (**RahmaniEtAl2015**).

Can this stark variation in prosodic type between MA and other dialects of Arabic be attributed to contact-induced change? The Arabic language has been in sustained contact with Amazigh (Moroccan Berber, MB) since the 7th century, but also with Latin, French and Spanish (Heath, this volume). **MaasProcházka2012** argue from corpus data that MA and MB share a common phonology, across a

range of segmental and suprasegmental features. **Bruggeman2018** confirms that there is no difference between MA and Tashlhiyt MB: both lack acoustic cues to word-level prominence in production and both groups of listeners display stress 'deafness'.

Since French is also an edge-marking language, without lexical stress, can we rule out French as an alternative source of this prosodic feature of MA? The main evidence comes from the fact that MA and MB also share other prosodic features which are not found in French, such as the shape of the tonal contour used to mark the edges of phrases, which is a rise in French (Delais-**RoussarieEtAl2015**), but a rise-fall in both Tashlhiyt MB (**GriceEtAl2015a**; **BruggemanEtAl2017**) and MA (**Benkirane1998**; Hellmuth to appear). The contrast is also exemplified in **Fagyal2005**'s (**Fagyal2005**) study of French-MA bilinguals in Paris who use an MA rise-fall contour in French.

## 2.3 Egyptian Arabic accent distribution

Cairene Egyptian Arabic displays a rich distribution of sentence accents, with a pitch accent typically observed on every content word. This has been noted independently by different authors (**Rifaat1991**; Rastegar-El **Zarka1997**), and is observed in both read and spontaneous speech styles (**Hellmuth2006**). Initial studies suggest that the same may be true also of some other dialects, such as Emirati (**BlodgettEtAl2007**) or Hijazi (**Alzaidi2014**), but these observations await corroboration across different speech styles.

Dense accent distribution has been noted in some languages on the northern coast of the Mediterranean also, including Spanish and Greek (**Jun2005**), although, in Spanish, the rich accent distribution seen in 'laboratory' speech is reduced in spontaneous speech (**Face2003**). Portuguese dialects vary in accent distribution: most varieties typically have an accent on every content word, but Standard European Portuguese (SEP) shows an accent on the first and last words in an utterance only (**FrotaEtAl2015**).

Rich accent distribution is not observed in Moroccan Arabic (**Benkirane1998**), nor in Tunisian Arabic (Hellmuth to appear). If the EA accent distribution pattern were due to contact between EA and the southern European languages on the other side of the Mediterranean which share the tendency towards rich accent distribution, we might expect the pattern to be found all across North Africa.

There is strong documentary evidence from written sources of historical sustained multilingualism in Egypt. Greek arrived in Egypt in the fourth century BCE, serving as a formal administrative language alongside Egyptian for several centuries, reaching a state of "balanced societal bilingualism" in Greek and

Egyptian in the sixth and seventh centuries CE (**Papaconstantinou2010**). Egyptian evolved into Coptic, and its prestige continued to increase from the sixth century CE onwards. After the Arab conquest in the seventh century CE, Arabic began to take over from Greek as the language of administration, eventually replacing Coptic in daily use (**Papaconstantinou2012**).

Is it possible that Egyptian/Coptic or Greek is the source of the rich accent distribution observed in EA (and indeed in Romance languages in southern Europe)?

The distribution of full and long vowels in Coptic indicates that it had word-level prominence (**Peust1999**), but it is not possible to determine from written texts the nature or distribution of any tonal contours which may have been associated with prominent syllables. Anecdotal evidence suggests that the intonation patterns used in surviving liturgical forms of Coptic are very different from those in EA (**Peust1999**), though this difference may owe more to the liturgical setting than to properties of the languages in spoken form.

Ancient Greek is generally thought to have had a pitch accent system in which the primary marker of culminative accent in each word was pitch (**DevineStephens1985**). The Koine Greek dialect used in Egypt is thought to have lost pitch accent in favour of a stress accent system, however, by the fourth century BCE (**Benaissa2012**).

Support for the hypothesis that Greek is the original source of the rich accent distribution would come from a match between the historical spread of Koine Greek around the Mediterranean with the location of languages in which rich accent distribution is also found. This would predict that eastern varieties of Libyan (Cyrenaican) Arabic might also be found to display rich accent distribution. If rich accent distribution is confirmed in dialects of Arabic (such as Emirati or Hijazi) which did not have sustained contact with Greek, or with EA more recently, this would argue against Greek as the original source. Although Nubi (**Gussenhoven2006**) and Juba Arabic (**Nakao2013**) display hybrid properties between stress and lexical tone, the most likely explanation of their prosodic patterns is direct contact with local tonal languages. A potential endogenous trigger for development of rich accent distribution would be the absence of other forms of phonological marking of word domains, which are indeed somewhat reduced in EA, in comparison to other dialects (**Watson2002**), though the direction of causality of this correlation is not easily determined.

Accent distribution has only recently been added to the parameters of variation explored in work on prosody (**Hellmuth2007**), and thus included in descriptions of the intonation systems of languages (e.g. **FrotaPrieto2015**). As further descriptions emerge of more dialects of Arabic it will be important to include

documentation of accent distribution, across genres and speaking styles, in future research.

## 3 Conclusion

There is much that we do not yet know about variation in intonation in Arabic, which leaves scope for investigation of further potential cases of contact-induced prosodic change. One such case may be the Syrian Arabic utterance-final rising intonation, sometimes known as 'drawl', which is found in yes/no-questions but also across other utterance types (**Cowell1964**), and which is an identifiable feature of the Damascus dialect (**KulkEtAl2003**). Although the full geographical range of the pattern has not been investigated in detail, and may be diffused to other dialects in the Levant, this rising declarative intonation pattern stands out from most other Arabic dialects, and is thus another potential case of contact-induced change.

Another potential outlier pattern is the rise-fall intonation contour seen in yes/no-questions in Yemeni Arabic (YA) from San'aa (**Hellmuth2014**). The full areal reach of this prosodic question marking strategy is also not yet fully known, and may extend into Hijazi Arabic and western dialects of Oman. However, we do know that a rise-fall is seen in both Tunisia and Morocco, though in these places the pattern may be due to contact with varieties of Amazigh Berber. Nevertheless it is tempting to speculate how a pattern found in Yemen might also be found in Tunisia and Morocco, and thus to explore the potential role of contact-induced variation due to ancient migrations between the eighth and fourteenth centuries (Holes in press).

Finally, the intonation of Modern Standard Arabic (MSA) and of other formal registers may prove to be a fruitful domain of future research. As our knowledge of the intonational phonology of spoken Arabic dialects improves, this will facilitate investigation of the extent to which the intonation patterns of a speaker's regional dialect can be observed and/or perceived in their MSA speech, building on observations in prior studies (El **ZarkaHellmuth2009**). An important goal would be to determine the extent to which a separate intonational system can be described for MSA, and to document the differential contribution to this system of specific genres of MSA discourse versus 'contact-induced' influence due to widespread community mastery of multiple registers of the language.

All these investigations would benefit from improved documentation of the time depth of present day surface intonation patterns. For the quasi-unique features explored in §**??**, we do not know whether these are the result of recent or

much more distant historical change. This situation might be rectified through analysis of archive audio materials, though dialect studies have often worked on oral narratives, which yield only a limited range of prosodic expression (i.e. usually few questions, and no information about turn-taking). A more viable strategy to gauge the time-depth of contact-induced variation in Arabic intonation would be for future sociolinguistic studies to include prosodic features as variables in 'apparent time' studies with participants in different age ranges, or for pre-existing corpora of apparent time data to be made available for prosodic analysis.

## Further reading

There are two key reference works, so far, on intonation in Arabic dialects, based on secondary analysis of prior published work: **Chahal2006** and El **Zarka2017**. Hellmuth (in press-a) suggests prosodic variables for inclusion in studies of variation and change in Arabic.

## Acknowledgements

## Abbreviations

EA Egyptian Arabic
   F0 fundamental frequency (which is the acoustic correlate of pitch)
   L1 first language, native language, mother tongue
   L2 second (or any additional) language
   MA Moroccan Arabic
   MB Moroccan Berber (Amazigh)
   TA Tunisian Arabic
   TB Tunisian Berber (Amazigh)
   MSA Modern Standard Arabic

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}
```

*??*

```
@misc{Alzaidi2014,
author = {Alzaidi, Muhammad},
note = {Unpublished PhD dissertation, University of Essex.},
title = {Information Structure and Intonation in Hijazi {Arabic}},
year = {2014}
}
```

```
@article{AttererAtterer2004,
author = {Atterer, M. and D. R. Ladd},
journal = {\textit{Journal} \textit{of} \textit{Phonetics}},
pages = {177--197},
title = {On the phonetics and phonology of "segmental anchoring" of F0: {{{E}}}v
volume = {32},
year = {2004}
}
```

Benaissa, Amin. 2012. Greek Language, Education, and Literary Culture (ed.), \te

Benkirane, T. 1998. Intonation in Western Arabic (Morocco). In D. Hirst \& A. D

```
@misc{BlodgettBlodgett2007,
author = {Blodgett, A., J. Owens and T. Rockwood},
note = {\textit{Proceedings} \textit{of} \textit{the} \textit{16th} \textit{ICPh
title = {An initial account of the intonation of Emirati {Arabic}},
urldate = {1137-1140},
year = {2007}
```

}

Bouchhioua, Nadia, Sam Hellmuth \& Rana Almbark. in press. Variation in prosodic a

Boudlal, A. 2001. Constraint interaction in the phonology and morphology of Casabl

@book{Bruggeman2018,
address = {Lexical and postlexical prominence in Tashlhiyt Berber and Moroccan Ara
author = {Bruggeman, Anna},
publisher = {Unpublished PhD thesis, Universität zu Köln},
year = {2018}
}

@book{BruggemanBruggeman2017,
address = {Question word intonation in Tashlhiyt Berber: is 'high' good enough? \te
author = {Bruggeman, Anna, Timo B. Roettger and Martine Grice},
publisher = {} \textit{Journal} \textit{of} \textit{the} \textit{Association} \tex
year = {2017}
}

@article{Bullock2009,
author = {Bullock, Barbara E.},
journal = {\textit{International} \textit{Journal} \textit{of} \textit{Bilingualis
pages = {165--194},
title = {Prosody in contact in {French}: {{A}} case study from a heritage variety

*??*

volume = {13},
year = {2009}
}

@article{BurdinBurdin2014,
author = {Burdin, Rachel Steindel, Sara Phillips-Bourass, Rory Turnbull, Murat Y
journal = {\textit{Lingua:} \textit{International} \textit{Review} \textit{of} \
pages = {254--276},
title = {Variation in the prosody of focus in head-and head/edge-prominence lang
volume = {165},
year = {2014}
}

@misc{CangemiCangemi2016,
author = {Cangemi, Francesco, Dina El Zarka, Simon Wehrle, Stefan Baumann and Ma
note = {\textit{Proceedings} \textit{of} \textit{Speech} \textit{\citealt{Prosod
title = {Speaker-specific intonational marking of narrow focus in {Egyptian} {Ar
year = {2016}
}

Cangemi, Francesco, Martine Grice \& Martina Krüger. 2015. Listener-specific per

Chahal, Dana. 2006. Intonation.  In K. Versteegh (ed.), \textit{Encyclopedia} \t

Chahal, Dana \& Sam Hellmuth. 2015. Comparing the Intonational Phonology of Leba

416

@article{ColantoniColantoni2004,
author = {Colantoni, Laura and Jorge Gurlekian},
journal = {\textit{Bilingualism:} \textit{Language} \textit{and} \textit{Cognition}
pages = {107--119},
title = {Convergence and intonation: {{H}}istorical evidence from {Buenos} {Aires}
volume = {7},
year = {2004}
}


@book{Cowell1964,
address = {\textit{A} \textit{reference} \textit{grammar} \textit{of} \textit{Syri
author = {Cowell, Mark W.},
publisher = {Georgetown University Press},
year = {1964}
}


D'Imperio, Mariapaola, Gorka Elordieta, Sonia Frota, Pilar Prieto \& Marina Vigari


Daoud, Mohamed. 2007. The Language Situation in Tunisia.  In R.B. Kaplan \& R.B.J.


Delais-Roussarie, E., B. Post, M. Avanzi, C. Buthke, A. Di Cristo, I. Feldhausen, S


@article{DevineDevine1985,

*??*

author = {Devine, AM and Laurence D Stephens},
journal = {Stress in greek? \textit{Transactions} \textit{of} \textit{the} \text
pages = {125--152},
volume = {115},
year = {1985}
}




@article{DupouxDupoux2001,
author = {Dupoux, Emmanuel, Sharon Peperkamp and Núria Sebastián-Gallés},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Acoustical} \textit
pages = {1606--1618},
sortname = {Dupoux, Emmanuel, Sharon Peperkamp and Nuria Sebastian-Galles},
title = {A robust method to study stress "deafness"},
volume = {110},
year = {2001}
}




@article{ElZarka2012,
author = {El Zarka, Dina},
journal = {Syntax, semantics and prosody of the Moroccan modal particle ra and i
pages = {412--431},
title = {{One} ra many meanings},
volume = {65},
year = {2012}
}




@misc{ElZarka2017,
author = {El Zarka, Dina},
note = {\textit{Oxford} \textit{Handbooks} \textit{Online}.},

```
title = {{Arabic} Intonation},
year = {2017}
}




@article{ElZarkaEl Zarka2009,
author = {El Zarka, Dina and Sam Hellmuth},
journal = {Variation in the intonation of Egyptian Formal and Colloquial Arabic  \
pages = {73--92},
volume = {22},
year = {2009}
}




@book{Face2003,
address = {Intonation in Spanish declaratives},
author = {Face, Tim},
note = {\textit{Catalan} \textit{Journal} \textit{of} \textit{Linguistics} 2  115-1
publisher = {differences between lab speech and spontaneous speech},
year = {2003}
}




@article{Fagyal2005,
author = {Fagyal, Zsuzsanna},
journal = {\textit{University} \textit{of} \textit{Pennsylvania} \textit{Working} \
pages = {91--104},
title = {Prosodic Consequences of Being a Beur: {{F}}rench in Contact with Immigran
volume = {10},
year = {2005}
}
```

*??*

Frota, Sonia, Marisa Cruz, Flaviane Fernandes-Svartman, Gisela Collischonn, Alin

```
@book{FrotaFrota2015,
address = {\textit{Intonation} \textit{in} \textit{Romance} Oxford},
author = {Frota, Sonia and Pilar Prieto},
publisher = {Oxford University Press},
year = {2015}
}
```

```
@misc{Gabsi2011,
author = {Gabsi, Zouhir},
note = {In \textit{International} \textit{Journal} \textit{of} \textit{the} \tex
title = {Attrition and maintenance of the Berber language in {Tunisia}},
year = {2011}
}
```

Gili Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Croc

```
@article{GriceGrice2015,
author = {Grice, Martine, Rachid Ridouane and Timo B. Roettger},
journal = {\textit{Phonology}},
pages = {241--266},
title = {Tonal association in Tashlhiyt Berber: {{{E}}}vidence from polar questi
volume = {32},
year = {2015a}
}
```

}

Grice, Martine, Michelina Savino, Alessandro Caffo \& Timo B. Roettger. 2015b. The

@misc{Gussenhoven2006,
author = {Gussenhoven, C.},
note = {\textit{Phonology} 23  192-223.},
title = {Between stress and tone in {Nubi} word prosody},
year = {2006}
}

@misc{Gussenhoven2017,
author = {Gussenhoven, C.},
note = {\textit{Journal} \textit{of} \textit{the} \textit{International} \textit{P[}
title = {Zwara (Zuwarah) Berber},
year = {2017}
}

Hansen, Anita Berit. 1997. Le nouveau [ə] prépausal dans le français parlé à Paris

Heath, Jeffrey. 2018. Moroccan Arabic.  In C. Lucas \& S. Manfredi (ed.), \textit{A

Hellmuth, Sam. 2006. Intonational pitch accent distribution in Egyptian Arabic: Un[

```
@article{Hellmuth2007,
author = {Hellmuth, Sam},
journal = {\textit{The} \textit{Linguistic} \textit{Review}},
pages = {289--314},
title = {The relationship between prosodic structure and pitch accent distributi
volume = {24},
year = {2007}
}
```

Hellmuth, Sam. 2014. Dialectal variation in Arabic intonation: motivations for a

Hellmuth, Sam. 2016. Explorations at the syntax-phonology interface in Arabic.

Hellmuth, Sam. in press-a. Prosodic Variation in Arabic.  In E. Al-Wer \& U. Hor

Hellmuth, Sam. in press-b. Text-tune alignment in Tunisian Arabic yes-no questio

Hellmuth, Sam. to appear. \textit{Intonation} \textit{in} \textit{spoken} \texti

```
@misc{HellmuthHellmuth2017,
author = {Hellmuth, Sam and Rana Almbark},
note = {UK Data Archive \url{http://reshare.ukdataservice.ac.uk/852878/}.},
```

title = {\textit{Intonational} \textit{Variation} \textit{in} \textit{{Arabic}} \te
year = {2017}
}



Herin, Bruno \& Martin R. Zammit. 2017. Three for the price of one: the dialects of



Holes, Clive. in press. Introduction (ed.), \textit{Arabic} \textit{Historical} \te



Jun, Sun-Ah. 2005. Prosodic typology.  In S.-A. Jun (ed.), \textit{Prosodic} \texti



Jun, Sun-Ah. 2014. Prosodic typology: by prominence type, word prosody, and macro-



Jun, Sun-Ah \& Janet Fletcher. 2014. Methodology of studying intonation: From data



@book{KulkKulk2003,
address = {The intonation of colloquial Damascene Arabic},
author = {Kulk, F., C. Odé and M. Woidich},
note = {\textit{IFA} \textit{Proceedings} 25  15-20.},
publisher = {a pilot study},
sortname = {Kulk, F., C. Ode and M. Woidich},
year = {2003}
}

*??*

```
@book{Ladd2006,
address = {Segmental anchoring of pitch movements},
author = {Ladd, D. Robert},
publisher = {autosegmental phonology or gestural coordination? \textit{Rivista}
year = {2006}
}
```

```
@book{Ladd2008,
address = {\textit{Intonational} \textit{phonology}  Cambridge},
author = {Ladd, D. Robert},
publisher = {Cambridge University Press},
year = {2008}
}
```

```
@article{LaddLadd1997,
author = {Ladd, D. Robert and R. Morton},
journal = {The perception of intonational emphasis: continuous or categorical? \
pages = {313--342},
volume = {25},
year = {1997}
}
```

Lucas, Christopher. 2014. Contact-induced language change.  In C. Bowern \& B. E

Maas, Utz. 2013. Die marokkanische Akzentuierung.  In R. Kuty, U. Seeger \& S. T

```
@article{MaasMaas2012,
author = {Maas, Utz and Stephan Procházka},
journal = {\textit{STUF-Language} \textit{Typology} \textit{and} \textit{Universal
pages = {329--357},
sortname = {Maas, Utz and Stephan Prochazka},
title = {Moroccan {Arabic} in its wider linguistic and social contexts},
volume = {65},
year = {2012}
}
```

Matras, Yaron. 2007. The borrowability of structural categories.  In Y. Matras \& .

```
@article{Mennen2004,
author = {Mennen, Ineke},
journal = {\textit{Journal} \textit{of} \textit{Phonetics}},
pages = {543--563},
title = {Bi-directional interference in the intonation of {Dutch} speakers of {Gre
volume = {32},
year = {2004}
}
```

```
@misc{Mitchell1993,
author = {Mitchell, T. F.},
title = {\textit{Pronouncing} \textit{{Arabic}} \textit{2} {Oxford}  Clarendon Pres
year = {1993}
}
```

Nakao, Shuichiro. 2013. The prosody of Juba Arabic: split prosody, morphophonolo

@article{Nance2015,
author = {Nance, Claire},
journal = {\textit{Lingua}},
pages = {1--19},
title = {Intonational variation and change in {Scottish} {Gaelic}},
volume = {160},
year = {2015}
}

O'Rourke, Erin. 2004. Peak placement in two regional varieties of Peruvian Spani

Papaconstantinou, Arietta. 2010. Introduction.  In A. Papaconstantinou (ed.), \t

Papaconstantinou, Arietta. 2012. Why did Coptic fail where Aramaic succeeded? Li

@book{Peust1999,
address = {\textit{Egyptian} \textit{phonology:} \textit{an} \textit{introductio
author = {Peust, Carsten},
publisher = {Peust und Gutschmidt},
year = {1999}
}

@article{Queen2012,
author = {Queen, Robin},
journal = {\textit{Language}},
pages = {791--816},
title = {{Turkish}-{German} bilinguals and their intonation: {{T}}riangulating evi
volume = {88},
year = {2012}
}


@article{Queen2001,
author = {Queen, Robin M.},
journal = {\textit{Language} \textit{in} \textit{Society}},
pages = {55--80},
title = {Bilingual intonation patterns: {{{E}}}vidence of language change from {Tu
volume = {30},
year = {2001}
}


@misc{RahmaniRahmani2015,
author = {Rahmani, Hamed, Toni Rietveld and Carlos Gussenhoven},
note = {\textit{PloS} \textit{one} 10.e0143968.},
title = {Stress "deafness" reveals absence of lexical marking of stress or tone in
year = {2015}
}


@misc{Rastegar-ElZarka1997,
author = {Rastegar-El Zarka, Dina},

```
note = {Unpublished PhD thesis, Karl-Franzens-Universität, Graz},
title = {\textit{Prosodische} \textit{Phonologie} \textit{des} \textit{Arabische
year = {1997}
}




@book{Rifaat1991,
address = {The intonation of Arabic: an experimental study},
author = {Rifaat, Khalid},
publisher = {Unpublished PhD thesis, University of Alexandria},
year = {1991}
}




@article{Sayahi2011,
author = {Sayahi, Lotfi},
journal = {Current perspectives on Tunisian sociolinguistics. \textit{Internatio
pages = {1--8},
title = {Introduction},
volume = {2011},
year = {2011}
}




@article{Simonet2011,
author = {Simonet, Miquel},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{International} \tex
pages = {157--184},
title = {Intonational convergence in language contact: {{{U}}}tterance-final F0
{Spanish} early bilinguals},
volume = {41},
year = {2011}
```

```
}




@article{Sorace2004,
author = {Sorace, Antonella},
journal = {\textit{Bilingualism:} \textit{Language} \textit{and} \textit{Cognition}
pages = {143--145},
title = {Native language attrition and developmental instability at the syntax-dis
volume = {7},
year = {2004}
}




@book{ThomasonThomason1992,
address = {\textit{Language} \textit{contact,} \textit{creolization,} \textit{and}
author = {Thomason, Sarah Grey and Terrence Kaufman},
publisher = {University of California Press},
year = {1992}
}




@article{Walker2014,
author = {Walker, Traci},
journal = {\textit{Research} \textit{on} \textit{Language} \textit{and} \textit{So
pages = {1--16},
title = {Form (does not equal) function: {{T}}he independence of prosody and actio
volume = {47},
year = {2014}
}
```

*??*

@book{Walters2011,
address = {Gendering French in Tunisia},
author = {Walters, Keith},
note = {In \textit{International} \textit{Journal} \textit{of} \textit{the} \tex
publisher = {language ideologies and nationalism},
year = {2011}
}

@book{Watson2002,
address = {\textit{The} \textit{phonology} \textit{and} \textit{morphology} \tex
author = {Watson, Janet C. E.},
publisher = {Oxford University Press},
year = {2002}
}

Watson, Janet C. E. 2011. Word stress in Arabic.  In M. Oostendorp, C. Ewen, E. 

Weinreich, Uriel, William Labov \& Marvin Herzog. 1968. Empirical foundations fo

@book{Wells1982,
address = {\textit{Accents} \textit{of} \textit{English} \textit{(Vols.} \textit
author = {Wells, John C.},
publisher = {Cambridge University Press},
year = {1982}
}

# Chapter 15

# Dialect contact and phonological change

??

??

This chapter examines phonological and phonetic changes that have been documented and analyzed in spoken Arabic varieties, occurring as a result of dialect contact. The factors contributing to dialect contact in Arabic speaking communities vary, from economic migration which has encouraged individuals to move into new dialect areas seeking work, to migration that stems from political violence and upheaval. These diverse factors have contributed to the large-scale migration of Arabic speakers to other parts of the Arabic speaking world. As a result, dialect contact is rampant, and decades of Arabic sociolinguistic research has shown that the phonological and phonetic outcomes of these contact situations have been quite profound.

## 1 Introduction

In this chapter, I discuss research that has examined the outcomes of Arabic dialect contact and the influence of contact on phonological change in spoken Arabic varieties. This chapter also discusses the interface between phonology and phonetics, and the effect of contact on these areas of the linguistic system. Given space constraints, I discuss only a portion of the published work in these areas, giving some priority to recent doctoral dissertations that have contributed to this body of research. Further, I exclude work that has investigated the effects of contact on the morphology and syntax of Arabic (e.g. Al-**WerEtAl2015**;**GafterHoresh2015**; Leddy-Cecere, this volume, Lucas, this volume, Mandfredi, this volume.

Although Arabic sociolinguistics is an increasingly robust area of linguistic research, limiting my discussion to cases of contact-induced phonological and phonetic change is perhaps unsurprising given the scholarly history of dialect

contact research and its place within sociolinguistics. Sociolinguistics has made great progress towards the goal of analyzing the full scope of variation in languages around the world. However historically, and to some extent still today, examinations of variation and change in the realms of phonology and phonetics have been the meat-and-potatoes of sociolinguistic work. I would argue that this is true of Arabic sociolinguistic work as well.

From **Labov1963**'s (**Labov1963**) early work on Martha's Vineyard, phonetics and phonology have been at the heart of analyses of dialect contact. As a result, much of what we know about Arabic dialect contact has stemmed from earlier foundational research on dialect contact in the English-speaking world. Within this work on English, research by **Milroy1987**, Trudgill (1986, 2004), **Britain2002**, and Britain and Trudgill (**BritainTrudgill2009**), among many others, has shown how dialect contact often plays out, and how that contact influences language variation and change.

However, research on Arabic has moved beyond simply testing the hypotheses put forward by scholars of English dialect contact, playing its own role in refining sociolinguistic theory. Notably, Arabic sociolinguistics has refined our understanding of diglossia (**Ferguson1959**). **Ibrahim1986** and **Haeri2000** have reoriented our understanding of Arabic diglossia from Ferguson's High–Low dichotomy to one that draws on locally meaningful understandings of linguistic prestige. In doing so, this work has moved our discussion away from analyzing Arabic through the lens of "standard" or "nonstandard" varieties or variants, setting the stage for decades of research that has examined contact-induced change in Arabic varieties.

## 1.1 The potential limitations of borrowing and imposition for Arabic dialect contact research

Before moving on to a discussion of a number of specific cases of Arabic dialect contact, I briefly address the potential limitations of Van Coetsem's (1988, 2000) framework for discussions of dialect contact, as opposed to language contact. After discussing Van Coetsem's approach, I shift my focus to discuss Arabic dialect contact through a theoretical lens that has proven productive in earlier sociolinguistic work (**Trudgill1986**, 2004).

In analyzing phonological change as a result of dialect contact, Van Coetsem's framework presents a number of possible challenges. One specific issue is that in many cases, a clear distinction between the borrowing or imposition of linguistic forms is challenging to establish in cases of Arabic dialect contact. Scholars may encounter challenges in attempting to assert the agentivity of the recipient lan-

guage in making a case for the borrowing of, for example, aspects of a dialect's phonology into the phonology of another dialect. Asserting the agentivity of the source language in making the case for imposition is similarly challenging. These challenges stem from the cognitive orientation of Van Coetsem's framework, which, as **Lucas2015** notes, is not based on social realities or variation in the power and prestige that a given dialect or language may hold.

The approach that many scholars within sociolinguistics and allied fields like linguistic anthropology have taken is, in contrast, inherently social. We concern ourselves with the social life of language, and although we do not discredit cognitive approaches to language acquisition and use, in much of the work on dialect contact, we have foregrounded social factors in our analyses of language change. However, it is worth noting that within sociolinguistic research on second dialect acquisition, researchers have highlighted the role of social factors in acquisition, as well as the constraints placed on acquisition by the linguistic system (e.g. **Nycz2013**, 2016).

With the above discussion in mind, I argue that Van Coetsem's framework is less readily applicable to the cases that I describe in this chapter. Instead, I suggest that outcomes of Arabic dialect contact are better analyzed through the framework advocated for within sociolinguistics. It is to that framework that I now turn.

As **Trudgill2004** notes in discussing new dialect formation, dialect contact often progresses in stages. One of the earliest stages in this process is leveling (**Trudgill2004**), which results in the reduction of forms from a given dialect. These forms may be, but do not have to be, socially marked, e.g. affrication of /k/ to [tʃ] in certain dialects. Most importantly for Trudgill, during leveling certain variants of a given feature will supplant others (**Trudgill2004**). As a result, forms that are socially marked may be leveled out, while unmarked forms may survive even if they were not a majority variant. In those cases where socially marked forms are present, they are often reduced across generations. Trudgill also describes processes of interdialect development, where forms arise out of the interaction between dialects, such as reallocation, where surviving forms are reallocated in some way, and focusing, whereby a new variety born out of contact begins to stabilize (**Trudgill2004**).

What I feel that this framework offers in discussions of Arabic dialect contact is an acknowledgement of the social issues that may influence linguistic change, especially in situations of contact. In the remainder of this chapter I discuss cases of contact-induced change in Arabic varieties. In doing so, I draw on sociolinguistic understandings of how contact-induced changes take hold and progress.

## 2 Contact-induced changes in the phonology of Arabic dialects

When discussing Arabic dialect contact, a brief discussion of the typology of Arabic is useful as it provides a shared lexicon for discussing the outcomes of contact. **Cadora1992** offers an ecolinguistic taxonomy of Arabic, describing a continuum of Arabic varieties containing linguistic features ranging from what he describes as Bedouin in provenance, to those that can be considered sedentary. In presenting a related contrast, Cadora describes features that situate dialects as being urban versus those that are rural.

However, what Cadora offers is not a hard and fast classification of Arabic varieties. Instead, his typology highlights linguistic features that typically group together within dialect types, providing a way to conceptualize the similarities and differences across these varieties. Important for this chapter, sites of contact between Arabic dialects are also often sites of contact between *types* of dialects as well.

In my own work on Palestinian Arabic this has been the case, with Gaza City offering one example of contact between different Palestinian Arabic dialects. Today, the dialect of Gaza City has both Bedouin and urban-sedentary features, dialect types that likely came into contact in Gaza as a result of Palestinian refugee migration (see de Jong's 2000 discussion of Gaza City). This contact is undeniable given Gaza's current demographic reality, which suggests that its population is roughly 70% refugee.[1] It is also unsurprising given that Gaza has long been a site of contact. This history of contact has resulted in a city dialect that looks different than other urban Palestinian varieties spoken in major cities like Jerusalem or Nablus.

The above example serves as a way to frame the linguistic discussion of contact-induced phonological change provided below. I begin by covering documented consonantal changes that have grown out of contact, before moving on to vocalic changes and the need for additional research in this area as studies of Arabic contact move forward.

---

[1]This figure has been reported by the United Nations Relief Works Agency for Palestinian Refugees but only reflects refugees that have actually registered with the U.N. Other estimates place the percentage of Gaza's population that are refugees as closer to 80%. https://www.unrwa.org/where-we-work/gaza-strip

## 2.1 Consonantal changes

One of the most widely discussed linguistic features within work on Arabic dialect contact has been the variable realization of the voiceless uvular stop /q/. Motivation for the scholarly interest in /q/ likely stems from a number of factors. First, the phoneme has a wide range of dialectal variation, with dialectal realizations including a true voiceless uvular [q], as well as [k, g, ʔ] and an additional [k] variant articulated between a velar and uvular (**Shahin2007**). Second, interest in /q/ is also likely due to the high social salience of its variation in many Arabic-speaking communities (see **Hachimi2012**; **CotterHoresh2015**).

The result is that /q/ has been one of the most heavily studied features in Arabic sociolinguistics. Variation and change in /q/ has been discussed in a number of different communities throughout the Arabic speaking world, including: Palestine (Abd El-**Jawad1987**; Al-**Shareef2002**; **CotterHoresh2015**; **Cotter2016**); Egypt (**Haeri1997**); Iraq (**Blanc1964**; Abu-**Haidar1991**);[2] Jordan (Abd El-**Jawad1981**; Al-**Wer2007b**; Al-**WerHerin2011**); Morocco (**Hachimi2007**, 2012); and Bahrain (**Holes1987**), among others.

What these cases suggest are robust processes of linguistic change in the realization of /q/ coming as a result of factors such as migration and dialect contact. While the social patterning of these changes (e.g. stratified along age, gender, or sectarian lines) have been as diverse as the communities in which /q/ has been analyzed, across these contexts we see regular patterns of change in /q/ over time.

Taking the case investigated by **Cotter2016** as an example, we can see how patterns of change in /q/ may progress over time. In the speech of Jaffa Palestinian refugees in the Gaza Strip, **Cotter2016** showed that across three generations Jaffa refugees in Gaza showed progressively lower use of their traditional [ʔ] of /q/, instead beginning to favor the voiced velar [g] variant that is common in Gaza City Arabic. Within the oldest generation of this community, Jaffa refugees showed near categorical retention of the glottal variant, and little rudimentary leveling. However, the second generation of Jaffa refugees showed substantial variability between [ʔ] and [g], while in the third generation in the study speakers showed higher rates of usage of the [g] variant that is native to the Gaza City dialect.

However, as **CotterHoresh2015** discuss, variability in /q/ is often is situated within broader identity projects that speakers and communities have undertaken. It is important then that analyses of Arabic dialect contact also consider the

---

[2]Both Abu-Haidar and Blanc's analyses are dialectological and descriptive in scope, however both ultimately discuss what appear to be processes of change taking place for /q/ within what Blanc termed "communal" (i.e. religio-sectarian) varieties of Arabic in Baghdad.

broader ethnographic context in which this contact takes place.

Another area of interest for researchers examining dialect contact has been the interdental fricatives, /θ, ð, ḏ̣/. Across Arabic varieties, these phonemes quite often vary between realizations as true interdental fricatives [θ, ð, ḏ̣] and their stop counterparts [t, d, ḍ] (Al-**Wer1997**, 2003, 2007a). In addition to descriptive work that has documented the realization of the interdentals across Arabic varieties, they have also been examined as sociolinguistic variables in cases of dialect contact.

For example, Holes (1987;1995) investigated sociolinguistic variation in the realization of /θ, ð, ḏ̣/ roughly split along sectarian lines in the speech of Arab and Baḥārna speakers in Bahrain. In Bahrain, in the dialect of Sunni Arab's these phonemes are traditionally pronounced as [θ, ð, ḏ̣], whereas in the dialect of Shi'i speakers they are pronounced as [f, d, ḍ]. Hole's (1995: 275) details that in the speech of young literate speakers in Manama, intercommunal dialect realizations of the interdentals have emerged that are generally centered on the Sunni Arab realizations of these phonemes. More recently, Al-**Essa2008** examined the interdentals in the speech of Najdi Arabic speakers living in Jeddah, Saudi Arabia, an Urban Hijazi Arabic dialect area. Although Najdi Arabic typically retains the interdental realization of these phonemes, Al-Essa concluded that degree of contact with Urban Hijazi speakers was a significant factor influencing whether Najdi speakers adopted the stop realizations common in Urban Hijazi Arabic.

Additionally, **Alghamdi2014** investigated the interdentals through the lens of migration and contact in the Saudi Arabian city of Mecca. Alghamdi describes what may be the beginning of a change from the traditional interdental realization of these phonemes in the direction of their stop counterparts. As Alghamdi notes (2014: 112), if it is the case that an incipient change in the interdentals exists in Mecca, the results of her study suggest that female speakers may be leading this change. This finding supports earlier sociolinguistic work, which has highlighted that female speakers are often at the vanguard of linguistic change.

Change in the interdentals has also been examined as part of new dialect formation in the Jordanian capital of Amman. As Al-**Wer2007b** describes (see also Al-Wer, this volume), Ammani Arabic has grown out of contact between speakers of two different dialect types: urban Palestinian and traditional Jordanian varieties, which differ in their realizations of the interdentals. Urban Palestinian dialects typically favor non-interdental realizations [t, d, ḍ], while, in contrast, traditional Jordanian dialects retain the interdentals [θ, ð, ḏ̣]. Al-Wer describes the case of the interdentals in Amman as a process of focusing (**Trudgill2004**) that has arisen out of contact. In Trudgill's terms (drawing on Le Page and Tabouret-

**Keller1985**), focusing is one part of the process of new-dialect formation, whereby features of input dialects are leveled and stability emerges resulting in new shared linguistic norms. Al-Wer describes that in Amman, focusing of the interdentals in the direction of their stop counterparts, [t, d, ḍ] has taken place (Al-**Wer2007b**: 66). In addition, Al-Wer notes that as a result of contact, Ammani Arabic has also focused towards the common Palestinian [ʒ] realization of /ǧ/ at the expense of the traditional Jordanian [dʒ] (Al-**Wer2007b**: 66).

In addition to the work by Al-**Essa2008** and **Alghamdi2014** discussed above, a more recent case of contact-induced change in Saudi Arabia has been identified: the voiced lateral fricative /ɮˤ/ realization of ⊠ Al-Wer and Al-**Qahtani2016** investigate /ɮˤ/ as a variable in the dialect of Tihāmat Qaḥtān. What this work shows is that in the Tihāmat Qaḥtān variety, the lateral [ɮˤ] represents a conservative, traditional variant of the phoneme, whereas the voiced interdental fricative [ð] represents the innovative variant. As a result of dialect contact, Al-Wer and Al-**Qahtani2016** note describe an intergenerational process of change towards the voiced emphatic interdental [ḍ], with use of the historic [ɮˤ] variant receding over time.

Another area of interest in dialect contact research has been affrication in Arabic dialects. As descriptive work has shown, affrication of certain phonemes, notably /k/ in the direction of [tʃ], is common in Arabic dialects. As an example of this process, **Shahin2007** notes that in rural varieties of Palestinian Arabic, /k/ palatalizes to become an affricate [tʃ] (e.g. [tʃīf-ak] 'how are you (m)?' < /kīf-ak/). While typologically this affrication is common, processes of affrication or de-affrication have also been noted as the outcome of contact.

Al-**Essa2008** investigated affrication of /k/ and /g/ in the speech of Najdi Arabic migrants in Jeddah, and found that the affrication that is a common feature of this variety had almost completely been undone in this migrant community. Examining this change in light of dialect contact, Alessa concludes that this deaffrication represents the leveling out of marked regional dialect forms as a result of contact (**Trudgill1986**; **KerswillWilliams2000**). More recently, Al-**WerEtAl2015** note that the conditional, root-based distribution of the affricate [tʃ] for /k/ in the as-Salt variety of Arabic in Jordan, which although it has receded (Al-**Wer1991**), now interacts with other innovative features in as-Salt that show potential stratification along religious lines.

Elsewhere in Jordan, notably in Amman, Al-**Wer2007b** describes the leveling of the affricate [tʃ] across generations. The city dialect that has emerged in Amman, which has Sulti Arabic as one of its input varieties, underwent rudimentary leveling (**Trudgill2004**) within the first generation. This leveling resulted in the

loss of this affricate variant of /k/ in the speech of Sulti migrants. In this case, Al-Wer describes the deaffrication of [tʃ] as stemming from its status as a marked feature of Horani Arabic varieties like that of as-Salt. This marked status makes it a primary candidate for the kinds of leveling that sociolinguists have identified in other cases of contact.

## 2.2 Vocalic changes

In general, the Arabic vocalic system remains understudied within research on Arabic varieties. However, multiple cases of change linked to dialect contact have been identified. One of the most well studied cases of contact-induced vocalic change in Arabic is perhaps better thought of as a morphophonological change: the Arabic feminine gender marker. The feminine gender marker is a word final vocalic morpheme that is realized variably across Arabic varieties. The realization of this vowel varies from an unraised [a] to [æ, ɛ, e], or even as high as [i] (e.g. Al-**Wer2007b**; **Naïm2007**; **Shahin2007**; **Woidich2007**).

Even within one region, the full range of variation in this morpheme can be seen. Taking the Levant as an example, the Lebanese capital of Beirut is known for raising this vowel to [e] or even [i] (**Naïm2007**). The Syrian capital of Damascus is known to raise to [e] (**Lentin2007**). Urban Palestinian (**Rosenhouse2007**; **Shahin2007**) is also often described as raising to [e], while Amman (Al-**Wer2007b**) raises this vowel to [ɛ]. These city varieties can be contrasted with, for instance, the variety of Cairo (**Woidich2007**), which does not raise this vowel, leaving it as [a].

This morpheme is particularly interesting within a discussion of dialect contact because raising of this vowel is phonologically conditioned. The phonological factors that constrain raising vary across dialects, with Urban Levantine Arabic (e.g. Syria, Palestine, Lebanon) providing one example of these factors. In Urban Levantine, the following rules constrain the raising of this vowel (**Grotzfeld1980**: 181; **Levin1994**: 44-45; Al-**Wer2007b**: 68):

1. The default realization of the vowel is raised, [e]

2. The vowel is unraised, realized as [a], when:

a. It occurs after back consonants (i.e. pharyngeal, glottal, emphatic/pharyngealized, post-velar) – ḥ, ʕ, ʔ, h, ṣ, ḍ/ḏ̣, ḫ, ɣ, q;

b. It occurs after /r/, but only when preceding /r/ there is no high front vowel. In cases where a high front vowel does precede /r/, raising is allowed, e.g. [kbiːre] 'big (f)'.

Below I provide two specific documented examples of contact and change in the feminine gender marker. First, **CotterHoresh2015** investigated change in the

feminine gender marker in the speech of refugees originally from the Palestinian city of Jaffa who now live as refugees in the Gaza Strip. This sample included both speakers who were expelled from Jaffa after the creation of the state of Israel in 1948 and their descendants. Their traditional Urban Palestinian dialect (**Horesh2000**; **Shahin2007**) is one that raises the feminine gender marker to [e], subject to the phonological conditioning mentioned above. In contrast, based on the available dialectological information the dialect of Gaza City does not raise this vowel (Bergsträßer 1915).

Cotter and Horesh (**CotterHoresh2015**) highlight a process of contact-induced change that has taken place in this community. Across generations, the realization of this vowel appears to be lowering and backing, moving from [e] in the direction of [a]. The result is that younger Jaffa refugee speakers realize the vowel closer to the [a] common in Gaza City. This type of change is perhaps unsurprising in a city like Gaza, given that the population of Gaza is overwhelmingly comprised of refugees, including large communities who are of [a] dialect types for this feature. This diversity and the high numbers of refugees in Gaza means that the city, and the territory generally, is a site where many dialects of Palestinian Arabic are in intimate contact. What remains to be determined is whether or not new linguistic norms are emerging in the dialect of Gaza City more generally as a result of this contact.

One other case, which is discussed in more detail in Al-Wer's contribution to this volume, provides a succinct example of the intersection between phonetics, phonology, and Arabic dialect contact. In discussing the formation of Ammani Arabic, Al-**Wer2007b** notes the centrality of vocalic change to the formation of the dialect. The feminine gender marker represents one feature that has helped to define the variety of Amman.

As Al-**Wer2007b** describes, through contact between Palestinian and Jordanian Arabic dialects in Amman, the realization of the feminine gender marker has focused on [ε], the indigenous Jordanian realization (see **Herin2014**). However, although Ammani Arabic has focused on the Jordanian phonetic realization of this vowel, it has retained Urban Palestinian phonology, which allows raising everywhere except after pharyngeal, velarized, and emphatic phonemes (Al-**Wer2007b**: 69). This is less restricted than the contexts in which raising is allowed in Horani Arabic, where raising occurs after the coronals /t,d,s,z,θ,.,ʃ, tʃ, dʒ, and n/, and after /b,m,f/ contingent on the presence of another front vowel (Al-**WerEtAl2015**: 77).

Finally, I mention one other case of vocalic change that has been documented as an outcome of dialect contact: the diphthongs [ay] and [aw]. **Alghamdi2014**

investigated monopthongization of the traditional Arabic diphthongs /ay/ and /aw/ in the speech of Nadji migrants in Mecca. Alghamdi found that the diphthongs common in the dialect spoken by this migrant community were monophthongizing, reflecting a change towards the norms of Mecca Arabic, which lacks diphthongs. Alghamdi's analysis of the diphthongs provides an example of dialect leveling borne out of contact, noting two additional aspects of this variable in the speech of this migrant community: i) Alghamdi describes the high degree of social salience that the diphthongs have in this community and their possible stigmatization in Mecca, and ii) that retention of the diphthongs is uncommon in Saudi Arabic, making the Ghamdi realization a minority realization in Saudi Arabia generally. These two facts create an environment conducive to change.

## 3 Conclusion

In examining Arabic dialect contact, a growing body of research highlights that the phonology and phonetics of Arabic represent rich sites for linguistic change. As the examples that I have provided throughout this chapter, and those discussed elsewhere throughout this volume suggest, we can identify a number of cases where dialect contact has influenced the directionality and extent of change in Arabic dialects. With the findings of this selection of work in mind, a number of areas remain open for future investigation.

Perhaps the most pressing of these is the reality that, although I have highlighted work here that investigates vocalic change, the vocalic system of Arabic varieties remains drastically understudied. Although phonetic research on the vocalic system of Arabic varieties continues to grow (see e.g. **HassanHeselwood2011**; Khattab and Al-**Tamimi2014**; Al-**TamimiKhattab2015**), we know little about sociophonetic changes that may take place in cases of contact like those discussed in this chapter. Given the scope of dialect contact in the Arabic speaking world, much of which has come as a result of mass migration throughout the region, investigating the potential for processes such as vocalic chain-shifting (Al-**Wer2007b**) represents an important next step for research on language variation and change in Arabic. I would argue that more robust investigation of vocalic change in Arabic dialects represents a pressing area of concern for Arabic sociolinguistics.

In addition, examples like the feminine gender marker in Amman (Al-**Wer2007b**) open the door for future work that investigates the potential for blending of the phonetics and phonology of different Arabic varieties as a result of contact. Although Amman represents a somewhat different case, given that it represents an example of new dialect formation, a close examination of phonetics and phonol-

ogy together in contact situations will provide us with an opportunity to examine how dialect focusing and leveling takes place, and how the linguistic systems of multiple different Arabic varieties interact and regularize through contact.

Additional research that looks more closely at change in the vocalic system of Arabic dialects will go a long way towards enriching the depth of Arabic sociolinguistic research. This is especially true of work that examines cases of dialect contact. However, beyond sociolinguistics, a closer examination of the vocalic system will contribute to the description and documentation of Arabic dialects, which will further enrich linguistic research that investigates the varieties of Arabic spoken around the world.

# Further reading

As I have discussed throughout this chapter, Al-**Wer2007b**'s (**Wer2007b**) work details a number of aspects of how the dialect of Jordan's capital, Amman, emerged as a result of contact between Jordanian and Palestinian dialects. It offers a clear picture of how contact has played out in Amman with respect to a number of linguistic features, and how the city's dialect emerged over successive generations.

Cotter and Horesh's ( 2015) article examines contact in the Gaza Strip, one of the more understudied areas within Arabic linguistic research. The article draws on sociolinguistic fieldwork conducted in Gaza City, as well Jaffa and the West Bank to analyze change in three specific features of Palestinian Arabic.

Clive **Holes1987**' (**Holes1987**) work on Bahrain provided one of the early accounts of sociolinguistic variation and change in Arabic. In addition, this work provides a clear example of variation in Arabic that has been stratified on sectarian, or religious lines.

# Acknowledgements

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}
```

```
@misc{AbdEl-Jawad1981,
author = {Abd El-Jawad, Hassan},
note = {" University of Pennsylvania.},
title = {"Lexical and Phonological Variation in Spoken {Arabic} in Amman},
year = {1981}
}
```

```
@article{———1987,
author = {———},
journal = {"Cross-Dialectal Variation in Arabic: Competing Prestigious Forms." \
pages = {359–67},
sortname = {———},
volume = {10},
year = {1987}
}
```

```
@book{Abu-Haidar1991,
address = {Wiesbaden},
author = {Abu-Haidar, Farida},
publisher = {Otto Harrassowitz},
title = {\textit{{Christian}} \textit{{Arabic}} \textit{of} \textit{Baghdad}},
year = {1991}
}
```

```
@book{Al-Essa2008,
address = {"Najdi Speakers in Hijaz},
author = {Al-Essa, Aziza},
note = {" University of Essex.},
```

```
publisher = {A Sociolinguistic Investigation of Dialect Contact in Jeddah},
year = {2008}
}




@book{Alghamdi2014,
address = {"A Sociolinguistic Study of Dialect Contact in Arabia },
author = {Alghamdi, Najla Manie},
note = {" Ph.D., University of Essex. http://ethos.bl.uk/OrderDetails.do?uin=uk.bl
publisher = {Ghamdi Immigrants in Mecca},
year = {2014}
}




@book{Al-Shareef2002,
address = {"Language Change and Variation in Palestine},
author = {Al-Shareef, Jamal},
note = {" University of Leeds.},
publisher = {A Case Study of Jabalia Refugee Camp},
year = {2002}
}




@book{Al-TamimiKhattab2015,
address = {"Acoustic Cue Weighting in the Singleton vs Geminate Contrast in Lebanes
author = {Al-Tamimi, Jalal, and Ghada Khattab},
note = {https://doi.org/10.1121/4.},
publisher = {344–60},
urldate = {1.492251},
year = {2015}
}
```

*??*

@book{Al-Wer1991,
address = {"Phonological Variation in the Speech of Women from Three Urban Areas
author = {Al-Wer, Enam},
publisher = {University of Essex},
year = {1991}
}


@book{———1997,
address = {"Arabic between Reality and Ideology." \textit{International} \textit
author = {———},
publisher = {251–65},
sortname = {———},
year = {1997}
}


@book{———2003,
address = {Versteegh, Martine Haak, and Rudolf De Jong, 21–31. Amsterdam},
author = {———},
publisher = {Brill},
sortname = {———},
title = {"Variability Reproduced: {{A}} Variationist View of the 'Daad'/Dhaa Opp
year = {2003}
}


@book{———2007,
address = {Leiden},

author = {———},
publisher = {Brill},
sortname = {———},
title = {"Phonological Merger." \textit{Encyclopedia} \textit{of} \textit{{Arabic}}
year = {2007a}
}




@book{———2007,
address = {E. Watson, 55–76. New York},
author = {———},
publisher = {Routledge},
sortname = {———},
title = {"The Formation of the Dialect of Amman: {{F}}rom Chaos to Order." In \text
year = {2007b}
}




@book{Al-WerAl-Qahtani2016,
address = {Amsterdam},
author = {Al-Wer, Enam, and Khairiah Al-Qahtani},
publisher = {John Benjamins},
title = {"Lateral Fricative Ḍād in Tihāmat Qaḥtān: {{A}} Quantitative Sociolinguist
69},
year = {2016}
}




@article{Al-WerHerin2011,
author = {Al-Wer, Enam, and Bruno Herin},
note = {. https://doi.org/10.3917/ls.138.0059.},
pages = {59–76},

*??*

title = {"The lifecycle of Qaf in {Jordan}." \textit{Langage} \textit{et} \texti
volume = {138},
year = {2011}
}




@article{Al-WerFanis2015,
author = {Al-Wer, Enam, Uri Horesh, Bruno Herin, and Maria Fanis},
journal = {"How Arabic Regional Features Become Sectarian Features? Jordan as a
pages = {68–87},
volume = {62},
year = {2015}
}




@book{Bergsträßer1915,
address = {Leipzig},
author = {Bergsträßer, Gotthelf},
publisher = {Hinrichs},
sortname = {Bergstrasser, Gotthelf},
title = {\textit{{Sprachatlas}} \textit{von} \textit{Syrien} \textit{Und} \texti
year = {1915}
}




@book{Blanc1964,
address = {Cambridge},
author = {Blanc, Haim},
publisher = {Harvard University Press},
title = {\textit{Communal} \textit{Dialects} \textit{in} \textit{Baghdad}},
year = {1964}
}

448

@book{Britain2002,
address = {"Diffusion, Levelling, Simplification and Reallocation in Past Tense BE
author = {Britain, David},
note = {https://doi.org/10.-9481.00175.},
publisher = {16—43},
urldate = {1111/1467},
year = {2002}
}


@book{BritainTrudgill2009,
address = {"New Dialect Formation and Contact-Induced Reallocation: Three Case Stu
author = {Britain, David, and Peter Trudgill},
publisher = {183—209},
year = {2009}
}


@book{Cadora1992,
address = {\textit{Bedouin,} \textit{Village,} \textit{and} \textit{Urban} \textit
author = {Cadora, F. J.},
note = {E.J. Brill.},
publisher = {} \textit{An} \textit{Ecolinguistic} \textit{Study}},
year = {1992}
}


@book{Coetsem1988,

*??*

address = {Dordrecht},
author = {Coetsem, Frans van},
publisher = {Foris},
title = {\textit{Loan} \textit{Phonology} \textit{and} \textit{the} \textit{Two}
year = {1988}
}




@book{———2000,
address = {Heidelberg},
author = {———},
publisher = {Winter},
sortname = {———},
title = {\textit{A} \textit{General} \textit{and} \textit{Unified} \textit{Theor
year = {2000}
}




@book{Cotter2016,
address = {Haddad and Potsdam, 229–46. Studies in Arabic Linguistics. Amsterdam}
author = {Cotter, William M.},
publisher = {John Benjamins},
title = {"(Q) as a Sociolinguistic Variable in the {{{Arabic}}} of Gaza City." I
year = {2016}
}




@book{CotterHoresh2015,
address = {"Social Integration and Dialect Divergence in~Coastal Palestine." \te
author = {Cotter, William M., and Uri Horesh},
note = {https://doi.org/10.1111/josl.12135.},
publisher = {460–83},

```
year = {2015}
}




@book{DeJong2000,
address = {Leiden},
author = {De Jong, Rudolf},
publisher = {Brill},
title = {\textit{A} \textit{Grammar} \textit{of} \textit{the} \textit{{Bedouin}} \t
year = {2000}
}




@article{Ferguson1959,
author = {Ferguson, Charles A.},
journal = {"Diglossia." \textit{Word}},
pages = {325–40},
volume = {15},
year = {1959}
}




@article{GafterHoresh2015,
author = {Gafter, Roey, and Uri Horesh},
journal = {"When the Construction Is Axla, Everything Is Axla: A Combined Lexical a
pages = {337–48},
volume = {3},
year = {2015}
}
```

*??*

@book{Grotzfeld1980,
address = {Wiesbaden},
author = {Grotzfeld, Heinz},
publisher = {Otto Harrassowitz},
title = {"Das Syrisch-Palästinesische Arabisch." In \textit{{Handbuch}} \textit{
90},
year = {1980}
}

@book{Hachimi2007,
address = {E. Watson, 97–122. New York},
author = {Hachimi, Atiqa},
publisher = {Routledge},
title = {"Becoming Casablancan: {{F}}essis in Casablanca as a Case Study." In \t
year = {2007}
}

@book{———2012,
address = {"The Urban and the Urbane: Identities, Language Ideologies, and Arabi
author = {———},
note = {https://doi.org/10.1017/S0040279.},
publisher = {321–41},
sortname = {———},
urldate = {740451200},
year = {2012}
}

@book{Haeri1997,

```
address = {New York},
author = {Haeri, Niloofar},
publisher = {Kegan Paul International},
title = {\textit{The} \textit{Sociolinguistic} \textit{Market} \textit{of} \textit{
year = {1997}
}




@book{———2000,
address = {"Form and Ideology: Arabic Sociolinguistics and Beyond." \textit{Annual
author = {———},
note = {https://doi.org/10.1146/annurev.anthro.29.1.61.},
publisher = {61–87},
sortname = {———},
year = {2000}
}




@book{HassanBarry Heselwood2011,
address = {Amsterdam},
author = {Hassan, Zeki Majeed, and Barry Heselwood, eds},
publisher = {John Benjamins},
title = {\textit{Instrumental} \textit{Studies} \textit{in} \textit{{Arabic}} \tex
year = {2011}
}




@book{Herin2014,
address = {Versteegh and Rudolf De Jong. \textit{Encyclopedia} \textit{of} \textit
author = {Herin, Bruno},
publisher = {Brill},
title = {"The Dialect of Sult." Edited by Kees},
```

*??*

year = {2014}
}


@book{Holes1987,
address = {London},
author = {Holes, Clive},
publisher = {Kegan Paul International},
title = {\textit{Language} \textit{Variation} \textit{and} \textit{Change} \text
year = {1987}
}


@book{———1995,
address = {"Community, Dialect and Urbanization in the Arabic-Speaking Middle Ea
author = {———},
publisher = {270–87},
sortname = {———},
year = {1995}
}


@book{Horesh2000,
address = {Paris},
author = {Horesh, Uri},
publisher = {AIDA},
title = {"Toward a Phonemic and Phonetic Assessment of Jaffa {Arabic}: {{I}}s It
20},
year = {2000}
}

@book{Ibrahim1986,
address = {"Standard and Prestige Language: A Problem in Arabic Sociolinguistics."
author = {Ibrahim, Muhammad H.},
publisher = {115–26},
year = {1986}
}




@article{KerswillWilliams2000,
author = {Kerswill, Paul, and Ann Williams},
journal = {"Creating a New Town Koine: Children and Language Change in Milton Keyn
pages = {65–115},
volume = {29},
year = {2000}
}




@book{KhattabAl-Tamimi2014,
address = {"Geminate Timing in Lebanese Arabic: The Relationship between Phonetic
author = {Khattab, Ghada, and Jalal Al-Tamimi},
note = {https://doi.org/10.1515/lp-2014-0009.},
publisher = {231–269},
year = {2014}
}




@article{Labov1963,
author = {Labov, William},
journal = {"The Social Motivation of a Sound Change." \textit{Word} 19 \REF{ex:key
},

*??*

url = {http://www.tandfonline.com/action/showCitFormats?doi=10.1080/00437956.196
volume = {309},
year = {1963}
}




@book{LePageTabouret-Keller1985,
address = {Cambridge},
author = {Le Page, Robert Brock Le, and Andrée Tabouret-Keller},
publisher = {Cambridge University Press},
sortname = {Le Page, Robert Brock Le, and Andree Tabouret-Keller},
title = {\textit{Acts} \textit{of} \textit{Identity:} \textit{{Creole}-Based} \t
year = {1985}
}




@book{Lentin2007,
address = {Versteegh. \textit{Encyclopedia} \textit{of} \textit{Arabic} \textit{
author = {Lentin, Jérôme},
publisher = {Brill},
sortname = {Lentin, Jerome},
title = {"Damascus {Arabic}." Edited by Kees},
year = {2007}
}




@book{Levin1994,
address = {Jerusalem},
author = {Levin, Aryeh},
publisher = {Magnes Press},
title = {\textit{A} \textit{Grammar} \textit{of} \textit{the} \textit{{Arabic}}
year = {1994}




456

}


@book{Lucas2015,
address = {London},
author = {Lucas, Chris},
publisher = {Routledge},
title = {"Contact-Induced Language Change." In \textit{The} \textit{{Routledge}} \
36},
year = {2015}
}


@book{Milroy1987,
address = {Oxford},
author = {Milroy, Lesley},
publisher = {Wiley},
title = {\textit{Language} \textit{and} \textit{Social} \textit{Networks}},
year = {1987}
}


@book{Naïm2007,
address = {Versteegh. \textit{Encyclopedia} \textit{of} \textit{Arabic} \textit{La
author = {Naïm, Samia},
publisher = {Brill},
sortname = {Naim, Samia},
title = {"{Beirut} {Arabic}." Edited by Kees},
year = {2007}
}

*??*

```
@book{Nycz2013,
address = {"Changing Words or Changing Rules? Second Dialect Acquisition and Pho
author = {Nycz, Jennifer},
note = {https://doi.org/10.1016/j.pragma.2012.12.014.},
publisher = {49–62},
year = {2013}
}
```

```
@book{———2016,
address = {Cambridge},
author = {———},
publisher = {Cambridge University Press},
sortname = {———},
title = {"Awareness and Acquisition of New Dialect Features." In \textit{Awarene
79},
year = {2016}
}
```

```
@book{Rosenhouse2007,
address = {Versteegh. \textit{Encyclopedia} \textit{of} \textit{Arabic} \textit{
author = {Rosenhouse, Judith},
publisher = {Brill},
title = {"Jerusalem {Arabic}." Edited by Kees},
year = {2007}
}
```

```
@book{Shahin2007,
```

```
address = {Versteegh. \textit{Encyclopedia} \textit{of} \textit{Arabic} \textit{La
author = {Shahin, Kimary},
publisher = {Brill},
title = {"Palestinian {Arabic}." Edited by Kees},
year = {2007}
}




@book{Trudgill1986,
address = {Oxford},
author = {Trudgill, Peter},
publisher = {Wiley-Blackwell},
title = {\textit{Dialects} \textit{in} \textit{Contact}},
year = {1986}
}




@book{———2004,
address = {Oxford},
author = {———},
publisher = {Oxford University Press},
sortname = {———},
title = {\textit{New-Dialect} \textit{Formation:} \textit{The} \textit{Inevitabili
year = {2004}
}




@book{Woidich2007,
address = {Versteegh. \textit{Encyclopedia} \textit{of} \textit{Arabic} \textit{La
author = {Woidich, Manfred},
publisher = {Brill},
title = {"Cairo {Arabic}." Edited by Kees},
```

*??*

```
year = {2007}
}
```

# Chapter 16

# Contact between Arabic and the Modern South Arabian Languages

??

??

> In the course of this chapter we will discuss what is known about the effects that contact with Arabic has had on the Modern South Arabian Languages (MSAL) of Oman and Yemen. Documentation concerning these languages is not abundant, and even more limited is our knowledge of the history of their interaction with Arabic. By integrating the existing bibliography with yet unpublished fieldwork materials, we will try to provide a picture of the situation as complete as possible, also discussing the current linguistic and sociolinguistic landscape of Dhofar and Eastern Yemen.

translations need single quotes

## 1 History of contact between Arabic and the Modern South Arabian Languages

Much to the frustration of modern scholars of Semitic, the history of the Modern South Arabian Languages (henceforth MSAL) remains largely unknown.[1] To this day, no written attestation of these varieties has been discovered, and it seems

---

[1] Paragraph

Is "section" intended??

1 was authored by Simone Bettega, while paragraph 2 by Fabio Gasparini. Paragraphs 3 and 4 are the result of the conjoined efforts of both authors. In particular, Gasparini was responsible for analyzing most of the primary sources and raw linguistic data, while Bettega worked more extensively on the existing bibliography.

safe to assume that they have remained exclusively spoken languages throughout all of their history. Since European researchers became aware of their existence, in the first half of the 19th century (**Wellsted1837**), and until very recently, the MSAL were thought by many to be the descendants of the old (epigraphic) South Arabian languages (**Rubin2014**). This assumption has been conclusively disproven by Porkhomovsky's 1997 article, which also contributed in a significant way to the re-shaping of the proposed model for the Semitic family tree. This modified version of the family tree (which finds further support in the recent works of Kogan, 2015 and **Edzard2017**) sets the MSAL apart as an independent branch of the West Semitic subgroup, one whose origins are therefore of considerable antiquity. This brings us to the question of when it was that the MSAL (or their forebears) first came into contact with Arabic. This might have happened at any time since Arabic-speaking people started to penetrate Southern Arabia, a process that – we know from historical records – began in the second half of the 1st millennium B.C. (**Robin1991**; **Hoyland2001**: 47-8). Roughly one thousand years later, almost the whole population of central and northern Yemen was speaking Arabic, and possibly a considerable portion of the southern population as well (**Beeston1981**: 184; **Zammit2009**: 295). It is therefore possible that Arabic and the MSAL have been in contact for quite some time, and it seems likely that the intensity and effects of such contact grew stronger after the advent of Islam (**Lonnet2009**). It is also possible, as some scholars have written, that the MSAL "represent isolated forms that were never touched by Arabic influence until the modern period" (**Versteegh2014**). Admittedly, evidence to support either one of these hypotheses is scarce, and at present it is probably safer to say that our knowledge of the history of contact between Arabic and the MSAL before the 20th century is fragmentary at best. This is why studies on the outcomes of such contact are of particular interest, since they could help shedding some light on parts of that history. This is also why, in the course of this chapter, we will refrain from addressing the question of how contact with the MSAL affected the varieties of Arabic spoken in Oman and Yemen, and focus solely on the influence of Arabic on the MSAL. Although there is plenty of evidence that South Arabian exerted a powerful influence on the Arabic of the area (see for instance **Retsö2000** and **Watson2017**)[2], it is often difficult to assess whether this influence is the result of contact with forms of Ancient South Arabian or more recent interaction with the MSAL. Such a discussion, also because of space constraints, is beyond the scope of the present article.

---

[2]To the point that "mixed varieties" are reported to exist, whose exact linguistic nature seems difficult to pinpoint. See **WatsonEtAl2006** and **Watson2011** for a discussion.

As far as the interaction between Arabic and the MSAL in the 20th and 21st century is concerned, **Morris2017** provides a good overview of the multilingual environment in which the MSAL were and are spoken:

> Speakers of an MSAL language always had to deal with speakers of other MSAL languages, as well as with speakers of various dialects of Arabic. The Baṭāḥirah, for instance, did nearly all their trade with boats from Ṣūr and other Arabic-speaking ports; they lived and worked beside the Arabic-speaking Janaba, and they were in contact with the MSAL speaking Ḥarāsīs and Mahra. The Ḥarāsīs interacted with the Arabic speakers surrounding their Jiddat al-Ḥarāsīs homeland, traded in the Arabic-speaking markets of the north, and in the summer months went to work at the northern date harvest. Mehri speakers lived beside and traded with Arabic-speaking Kathīri tribesmen in the Nejd region, Śḥerɛt speakers in the mountains, and Arabic speakers in the coastal market towns of Dhofar. Śḥerɛt speakers interacted with the Mahra, some of whom settled among them, and with Arabic-speaking peoples of the coast as well as the desert interior […] There was marriage between Arabic-speaking men of the coastal towns and MSAL -speaking women of the interior, and over time, families of Mehri- and Śḥerɛt-speakers settled in or near the towns, with the result that even more Arabic speakers became familiar with these languages.

## 2 Current state of contact between Arabic and the Modern South Arabian Languages

Today, six Modern South Arabian Languages exist, spoken by around 200,000 people in Eastern Yemen (including the island of Soqotra) and Western Oman. These six languages are: Mehri, Hobyōt, Ḥarsūsi, Baṭhari, Śḥerɛt/Jibbāli and Soqoṭri. They are all to be regarded as endangered varieties, though the individual degree of endangerment varies remarkably. No exact census concerning the number of speakers is currently available (Simeone-**Senelle2011**: 1075), but we know that Mehri is the most spoken language, with an estimate of around 100,000 speakers. It is followed by Soqoṭri (about 50,000 speakers), Śḥerɛt (25,000), Ḥarsūsi (??), Hobyōt (a few hundreds) and Baṭhari (less than 20 speakers). The main causes of endangerment are reckoned to be contact with Arabic and disappearance of traditional local lifestyles. In addition, the current political situation in Yemen is having effects on the linguistic landscape of the region which are difficult to document or foresee: the area is currently inaccessible to researchers, and

there is no way to know how the conflict will affect the local communities.

As far as Oman is concerned, the city of Salalah undoubtedly represents the major locus of contact between Arabic and the MSAL. The rapid growth the city has witnessed in recent years, and the improved possibilities of economic development that came with it, have led many Śḥerɛt-speaking breeders from the nearby mountains to settle down in the city or its immediate surroundings, where they now employ Arabic on a daily basis as a consequence of mass scholarization and media neglecting other local languages. This has led to a split, in the speakers' perception, between 'proper' Śḥerɛt, spoken in the mountains, and the 'city Śḥerɛt' of Salalah, often regarded as a sort of 'broken' variety of the language in which, among other things, code-switching with Arabic is extremely frequent. Unfortunately, data on this subject are virtually non-existent, given the extreme difficulty of documenting such an episodic phenomenon (aggravated by the speakers' understandable reluctance to having their shaky language proficiency evaluated and recorded).

Even outside the urban centers, however, contact with Arabic is on the rise. Even the most isolated variety, Soqotri, is apparently undergoing rapid change under the influence of Arabic: the existence of a "koineised" variety of Soqotri, heavily influenced by Arabic, has been recently reported in Ḥadibo (**Morris2017**). This is not to say, of course, that all MSAL are being affected to the same degree: **Watson2012**, for instance, notes how "Mahriyōt [the EasternYemeni variety of Mehri] [...] exhibits structures unattested in Mehreyyet [Mehri Omani variety] [...] and shows greater Arabic influence both in terms of the number of Arabic terms used, and the length and frequency of Arabic phrases within texts". However, no MSAL seems at present to be exempt from the effects of contact.

The case of Baṭḥari, in particular (which – as we have seen – is the most severely endangered of all the MSAL), exemplifies well the processes of morphological loss and erosion that a language undergoes in the final stages of endangerment. **Morris2017** reports how already in the 1970s "Baṭḥari seemed to display many of the signs of a moribund language". In recent times

> "[t]he younger generations showed little interest in their former language; they were eager to embrace Arabic and to feel themselves part of the wider Arabic Islamic community; and they were proud to call themselves 'ʕarab', with all that word's overtones of Bedouin ancestry and code of honour (**Morris2017**).

> In the following sections we will discuss several types of contact-induced changes in the MSAL: although we will use material taken from all varieties,

Baṭḥari will be in particular focus due to its singular status.

# 3 Contact-induced changes in the MSAL

As already anticipated, in the course of this chapter we will focus solely on the effects that contact with Arabic has had on the various MSAL. Therefore, Arabic will always be the Source Language of all the transfer phenomena considered in the next pages, while the Recipient Language will be, depending on the different examples, one or the other of the six MSAL. Obviously, this poses the question of who the agents of change are and have been in the case of these particular phenomena, and what type(s) of transfer are we confronted with. According to the overview of the MSAL's sociolinguistic status presented above, it should be clear by now that, while the two cases are extremely common of (a) mono- or multilingual MSAL speakers who acquire Arabic as an L2 and (b) bilingual MSAL-Arabic speakers, the opposite is not true (that is, monolingual Arabic speakers who come to acquire one or more MSAL as L2s later in life). In other words, all the transfer phenomena we will be considering in the next paragraphs are either instances of borrowing (brought about by speakers who are dominant in one or more MSAL) or convergence (brought about by speakers who are native speakers of Arabic and at least one MSAL, see **Lucas2015** for a definition of convergence).

## 3.1 Phonology

### 3.1.1 Phonetic adaptation of loanwords

As illustrated in §??, lexical borrowings from Arabic are extremely common in the MSAL. As **Morris2017** remarks, such loanwords are often altered in order for them to acquire a "South Arabian flavor", so to speak. The phenomenon is not one of simple adaptation dictated by difficulty of articulation, since the sounds that get replaced are present in the phonological inventory of the MSAL. In fact, the opposite appears to be true, these sounds being normally replaced by others which are typical of South Arabian but absent in Arabic. For Baṭḥari, Morris gives the example of Arabic pharyngealised dental fricative /ḍ/ (IPA [ðˤ]) being replaced by the pharyngealised alveolar lateral fricative /ṣ́/ (mostly realised as IPA [ɬˤ], see 3.1.4), as in *raṣ́ṣ́* 'bruise' (from Janaybī Arabic *raḍḍ*), or Arabic /š/ (IPA [ʃ]) being replaced by /ṣ́/, as in *men ṣ́ān-k,* 'for you, for your sake', in place of *men šān-k*, *ṣ́arrày*[3] 'buyer' for *šarrày*, or *ṣ́əmāl* 'inland, north' for *šəmāl* (while

---

[3]Through the course of this chapter we have used downward-slanting accents to indicate stress accent when falling on short vowels.

Baṭḥari *śēməl(i)* is normally used to refer to the left hand only).

Lexical borrowing can also be the cause of variation in the realisation of certain sounds, as is the case with the phonemes /g/ and /y/ (IPA [g] and [j] respectively), which represent possible realisations of an underlying /ǧ/ (IPA [ʤ]) in different Omani Arabic dialects. It is possible to find traces of this variation in those MSAL that are in contact with more than one dialect of Arabic, as is the case with Ḥarsusi: see for instance *fagr* and *fayr*, both meaning 'dawn', or the opposition between *yann* 'madness' and *gènni* 'jinni', both from the same etymological root (**Lonnet2009**).

### 3.1.2 Affrication of /k/ > [tʃ]

It can also be the case that some phonetic processes regularly taking place in the local Arabic varieties but otherwise unknown to MSAL phonology are transferred to original MSAL vocabulary. This is what happens in Baṭḥari, where some speakers may show an affricate realisation of the voiceless occlusive [k] > [tʃ], which resembles the Janaybī Arabic realisation of the phoneme /k/ (whose complementary distribution with the voiceless plosive realisation [k] is still unclear). For example, some speakers regularly produce /yənkàʕ/ 'to come.3S.M.SBJV' as [jən'tʃaʕ] instead of [jən'kaʕ].

### 3.1.3 Stress

The structural similarity of Arabic and the MSAL can sometimes cause stress patterns which are typical of the former to be applied to the latter, as is the case with "she began" Soqotri *bèdʔɔh*, (local) Arabic *bədàʔat,* Soqotri with an Arabic stress *bədɔʔɔh* (**Lonnet2009**).

### 3.1.4 Realisation of emphatics

This is a topic that has attracted the attention of several scholars since the publication of Johnstone's 1975 article on the subject, because of the realisation of the so-called Semitic "emphatics" as glottalised consonants. Glottalisation is a secondary articulatory process in which narrowing (creaky voice) or closure (ejective realisation) of the glottis takes place: the action of the larynx compresses the air in the vocal tract which, once released, produces a greater amplitude in the stop burst (**LadefogedMaddieson1996**: 78).

**Lonnet2009** notes a tendency for speakers of various MSAL to replace the ejective articulation of certain consonants (especially fricatives, see **RidouaneGendrot2017**)

with a pharyngealised realisation, typical of Arabic "emphatics". Pharyngealisation is a kind of secondary articulation involving a constriction of the pharynx usually realised through tongue root retraction, resulting in a "backed" realisation (**LadefogedMaddieson1996**: 365) and it is a well-documented process across Semitic languages. **NaumkinPorkhomovsky1981** note for Soqoṭri an ongoing process of transition from a glottalised to a pharyngealised realisation of "emphatics", only stops being realised as fully glottalised items. Work from Watson & Bellem (2010; 2011) and **WatsonHeselwood2016** show the co-occurrence of pharyngealisation and glottalisation in relation to pre-pausal phenomena in Sanaani Arabic, Mahryōt and Mehreyyet (respectively the westernmost Yemeni and Omani varieties of Mehri). **Dufour2016** states that "le caractère éjectif des phonèmes emphatiques ne fait aucun doute, en jibbali comme en mehri" ("the nature of the emphatic phonemes is undoubtedly ejective, in Jibbali as much as in Mehri"[4]). Finally, in Baṭḥari only /ḳ/ is realised as a fully ejective consonant [k']. The /ṭ/ and fricative emphatics, on the contrary, are described as mainly pharyngealised (and partially voiced, in the case of fricatives; **Gasparini2017a**).

Unfortunately, since no thorough phonetic description of any MSAL exists that predates the 1970s, it is impossible to ascertain whether these realisations (which, again, range from fully glottalised to fully pharyngealised) are the result of the influence of Arabic or have arisen as the consequence of internal and typologically predictable developments. It is likely, though, that bilingualism and constant contact with Arabic have at least favoured this phonetic change. Evidence in support of this view may come from the fact that speakers who are poorly proficient in Arabic and live in rural and more isolated areas are more likely to preserve a glottalised realisation of the emphatics (as emerges from direct fieldwork observations).

## 3.2 Morphology

### 3.2.1 Nominal morphology

Morphological patterns which are typical of Arabic can penetrate the language through borrowing, as is the case of the past participle of simple verbs which is *mvCCūC* in Arabic and *mvCCīC* in MSAL. Soqoṭri *màxlɔ̣ḳ*, for instance, is clearly derived from Arabic *maxlūq* 'human being' (lit. 'created'), while this is not the case for Ḥarsusi *mxəlīḳ* (**Lonnet2009**). Also, in the realm of verbal derivational morphology, certain phenomena can be introduced in the recipient language through lexical borrowing: it is the case of gemination and prefixation of *t-* in

---

[4]Our translation.

Ḥarsusi, as in the participle *məṭḥàffi* 'barefoot' (from Omani Arabic *miṭḥàffi*; **Lonnet2009**: 299).

In general, Arabic loanwords are normally well integrated in MSAL morphology, probably because of the high degree of structural similarity that exists between these languages. One example, reported by **Lonnet2009**, is that of *bəḳerēt* 'cow', a fully integrated loan from Arabic used in Ḥarsusi and Western Yemenite Mehri, which possesses its own plural and diminutive form (*bəḳār* and *bəḳərēnōt*, respectively).

Arabic loans in several MSAL stand out because of their characteristic feminine ending in *-v(h)* instead of *-(v)t*, as in Śḥehri *sáʕáh* 'watch' and *ṭòrəh* 'revolution' (but consider the more adapted *rišt* 'trigger' from Omani Arabic *rīšah*; **Lonnet2009**: 300).

It is also worth noting that the Arabic ending *-v(h)* is substituted by its MSAL equivalent when the noun is in the construct state, i.e. final *-t* reappears. This would also happen in Arabic, but the alteration in the quality of the vowel is a clear signal that the suffix is to be considered an MSAL morpheme. Consider the following example from Morris' Baṭḥari recordings:

(1)  Baṭḥari[5]

mʕayš-ìt=həm /bəss!   mʕayš-ìt=həm ḥawla   ʕār   ḥāmis
sustenance-F= POSS.3PL.M only       sustenance-F= POSS.3PL.M once

w=ṣayd /   śālā   mʕayš-ah ḥawīl
only    turtle and=fish nothing   sustenance-F once

"Their sustenance /only that! Their sustenance once was only turtle and fish / there was nothing to eat in the past".

The word *mʕayšah* 'sustenance, food' is a loanword from Arabic (as the *-ah* ending suggests). When suffixed with the possessive 3PL.M pronoun *=həm*, however, Baṭḥari *-ìt* replaces *-ah/-at* (note also, in the example, the use of the restrictive adverbial particle *bəss* 'only', which is a well-integrated loan from dialectal Arabic and occurs in alternation with Baṭḥari *ʕār*).

Finally, in Baṭḥari the Arabic definite article *(a)l=* is occasionally used instead of the MSAL definite article *a=*: *bə=l=xarifēt*, 'during the Xarif season'.

---

[5] Audio file '20130929_B_B02andB04_storyofcatchingturtle' recorded, transcribed and kindly shared with Fabio Gasparini by Miranda Morris. The recording was produced in the context of Morris' and Watson's "Documentation and Ethnolinguistic Analysis of the Modern South Arabian languages" project, funded by Leverhulme Trust. More recordings are accessible at the ELAR archive of SOAS University of London. The transcription has been adapted.

### 3.2.2 Pronouns

The influence of Arabic can be observed, to an extent, even in the pronominal system, especially in those MSAL that are more exposed to contact due to the limited size of their speech communities. **Lonnet2009**, for instance, reports how, despite the fact that in the MSAL the first person suffix pronoun is normally an invariable *=i*, in Ḥarsusi this can be replaced by *=ni* after verbs and prepositions (as is the case with Arabic; see also §**??** for another interesting example concerning the marking of pronominal direct objects).

In addition, Baṭḥari relative pronouns (S: *l-, lī* PL: *əllī*) are close to their equivalent in Janaybī Arabic (and diverge from the rest of MSAL, where a *ḏ=* element can be found). Baṭḥari has also borrowed the reflexive pronoun *ʕamr-* 'oneself' from the Arabic dialect of the Janaba, despite the existence of an original Baṭḥari term with the same meaning, *ḥanef-* (note that both terms must always be followed by a suffix personal pronoun). *ʕamr-* has also been given a plural form in Baṭḥari, based on MSAL derivational patterns, *ḥaʕmār-* (**Morris2017**).

### 3.2.3 Baṭḥari verbal plural marker *-uw*

Baṭḥari differs from the rest of the MSAL in that all 2/3PL.M verbal forms are marked by an *-uw* suffix, while in the other languages of the group these persons are marked by a *-əm* ending and/or by internal vowel change (i. e. Mehri and Ḥarsusi *-kə(u)m* for the 2PL.M and *-ə(u)m/*umlaut for the 3PL.M of the perfective conjugation*; t-…-ə(u)m* and *y/i-…-ə(u)m* respectively for 2 and 3PL.M of the imperfective conjugation; Simeone-**Senelle2011**: 1093-1094).

The origin of this suffix is uncertain. Its presence might well be connected to contact with Arabic (neighboring dialects have an *-u* or *-ūn* suffix in the 3PL.M person of the verb in both the perfective and imperfective conjugation) or to otherwise unattested stages of development internal to the MSAL verbal system. In this regard, **Rubin2017** suggests for Mehreyyet the presence of a subjacent *-ə-* in 2nd/3rd plural masculine verbs which could therefore be somehow related to the Baṭḥari *-uw* marker. However, the optional simultaneous presence of apophony within the stem of 3PL.M verbal forms (similarly to what happens elsewhere in the MSAL), together with scarcity of data, prevents any conclusive assessment of the topic.

## 3.3 Syntax

At present, the syntax of the various MSAL has not been made the object of de-
tailed investigation. The only scientific work dealing with this topic is **Watson2012**'s
(**Watson2012**) in-depth analysis of Mehri syntax. However, Watson's thorough
description provides only sporadic insights on the issue of language contact (as
for instance the use in Mahriyōt, the Eastern Yemeni variety of Mehri, of a *swē
~ amma... yā* construction to express polycoordination, probably to be regarded
as the result of Arabic influence; **Watson2012**: 298). In general, though, the topic
is left unaddressed in the literature, and more research is needed.

An interesting example of Arabic influence on MSAL syntax comes from Gas-
parini's data on Baṭḥari. In Baṭḥari, as in the other MSAL, pronominal direct and
indirect objects may require a particle *t-* to be inserted between them and the
verb, depending on the morphological form of the verb itself. Masculine singular
imperatives, for instance, require the presence of the marker, as shown in the
following example:

(2)  Baṭḥari (Gasparini's unublished data):
     zùm        t=ī     t=ih
     to_give.IMP ACC=1S ACC=3S.M
     "give it to me"

If we now compare (**??**) with (**??**):

(3)  Baṭḥari (**Gasparini2017b**):
     zɛm=ənī       t̠rɛh
     to_give.IMP=1S two.M
     "Give me both (of them)"

We see that the pronominal indirect object *=(ə)nī* is suffixed directly to the verb
as it would be in Arabic (see §**??**). In other words, the introduction of the Arabic
form of the object pronoun has caused the Baṭḥari object marker to disappear.
Note that informants judged the alternative construction *zùm t=ī t̠rɛh,* (with the
use of the object marker *t=* and the 1s object pronoun marker *=ī*) to be acceptable,
but this form was not produced spontaneously.

A peculiarity of the MSAL spoken in Oman is the use of circumstantial quali-
fiers, a type of clausal subordination well attested in Gulf Arabic (**Persson2009**).
Baṭḥari regularly introduces predictive and factual conditional clauses asyndeti-
cally by using the structure [SBJ.PN w=SBJ.PN]. Consider (**??**):

(4)  Baṭhari (Gasparini's unpublished data)

hēt  w=hēt    t-áṣbaḥ                 axayr /               saḥīr-e
2s.m and=2s.m 2s.m-to_wake_up.ipfv better to_brand.ptcp-s.m acc=2s.m

t=ōk lā   / w               ham aṣbaḥ-k      axáss /
neg  and if to_wake_up.pfv-2s.m worse to_hear.pfv-2s.m and  illness

hāmā-k? w       maráḏ zēd              l=ōk     / nḥā saḥīr-en t=ōk
huge    to=2s.m 1pl   to_brand.pfv-1pl acc=2s.m

If you woke up feeling better / I would not brand you / but in case you
woke up feeling worse / do you understand? And you were seriously ill /
we would brand you

The first clause / *hēt w=hēt t-áṣbaḥ axayr* / is an asyndetical circumstantial
qualifier functioning as a predictive conditional clause. It contrasts with / *w ham
aṣbaḥ-k axáss* /, in which the conjunction *ham* introduces a counterfactual con-
ditional clause.

In Omani Mehri conditional clauses are commonly introduced through con-
junction of pronouns (Watson forthcoming: 187). This structure is unattested in
Yemeni Mehri:

(5)  Mehri (Watson forthcoming: 188)

*sēh  wa=sēh   t-ḥam=ah                 lā   /*
3s.f and=3s.f 3.s.f-to_want.ipfv=3s.m neg parents=poss.3s.f

*ḥib=sa         yi-ḳal-am t=ēs          ta-ghōm   š=ih lā* If she
3m-to_let.ipfv-pl acc=3s.f  3s.f-to_go.sbjv with=3s.m neg
doesn't want him / her parents won't let her go with him

These uses closely resemble those of Gulf Arabic, where circumstantial quali-
fiers are widely attested to codify predictive and factual conditional and consec-
utive clauses.

## 3.4 Lexicon

In the case of the msal, it can often be difficult to clearly set apart the effects
of Arabisation from those of modernisation and lifestyle changes (which is not
surprising, since the two phenomena are interrelated). According to what the
speakers themselves report,

> it was only since the introduction of formal education, and the awareness
> of m[odern] s[tandard] a[rabic] via the media, that Arabic became the sec-
> ond language for many of the msal speakers in Dhofar, and, in the case of

younger speakers, often to the detriment of their proficiency in their MSAL variety (**Davey2016**).

As a consequence, phenomena of borrowing (such as codeswitching and loan-words) are particularly common, especially in those varieties and in the idiolect of individuals which are more exposed to Arabic. The following is a good example of codeswitching in Baṭḥari (note that the speaker in question tended to employ Janaybi forms more than other informants):

(6) Baṭḥari (Gasparini's unpublished data)
mɛ̄ t məssəlīm /
to_die.PFV.3S.M muslim 1PL-to_say_šahada.IPFV for=2S.M
nə-šāhəd l=ōk /
and=3M-to_pray.IPFV-PL and=3M-to_pray-PL and=3M-to_praise_allah-PL
w=y-sabbah-uw w=y-kabbər-uw w=y-hālul-uw

"(If) a muslim dies / we say the *šahada* for you / and they pray and say '*allahu akbar*' and praise Allah"

In (**??**) the speaker makes use of several Arabic verbs related to the semantic field of religious practices, which are not lexically encoded in Baṭḥari. This might indirectly show the introduction of new ritual practices at a certain point of the history of the tribe. Note that $c^2$-geminate stems such as *ysabbahuw* and *ykabbəruw* represent verbal patterns not attested in MSAL morphology, and are therefore easily identifiable as loans.

**Morris2017** makes the important remark that lexical erosion is directly connected with the loss of importance of a language in the eyes of its speakers. She gives the example of the Baṭḥari word for 'home, living quarters', for which speakers nowadays frequently resort to some version of Arabic *bayt*, while the many possible original synonyms are falling into disuse. Many of these (*kədōt*, *mōhèn*, *màšʕar*, *mōxayf*, and *xàder*) are connected to traditional ways of living which have all but disappeared in the course of the last 40-50 years, so that speakers probably judge them inadequate to refer to modern built houses.

### 3.4.1 Numbers

**Watson2012** reports that '[w]hile Mehri cardinal numbers are typically used for both lower and higher cardinals in Mehreyyet, Mahriyōt speakers, in common

with speakers of Western Yemeni Mehri, almost invariably use Arabic numerals for cardinals above 10'. This type of lexical substitution connected to numerals higher than ten is also mentioned by **Lonnet2009** and Simeone-**Senelle2011**, who states that '[n]owadays the MSAL number system above 10 is only known and used by elderly Bedouin speakers'. Watson and Al-**Mahri2017** note that it is mostly younger generations (especially in urban settings) who have lost the ability to count beyond ten. Interestingly, they point out how telephone numbers are given exclusively in Arabic, "possibly due to the lack of a single-word MSAL equivalent to Arabic *ṣufr* 'nothing'".

### 3.4.2 Spatial reference terms

According to Watson & Al-**Mahri2017** the MSAL employ topographically variable absolute spatial reference terms. In other words, these terms can differ depending on the language employed, the moment of the utterance and the position of the speaker in relation to absolute points of reference. For instance, in and around the city of Salalah in Dhofar, both in the mountains and on the coastal plain, the equivalents of the words for 'sea' and 'desert' are used to indicate south and north, respectively, in both Mehri (*rawram* and *nagd*) and Śḥehri (*ramnam* and *fagir*). This is because the sea lies to the south and the desert lies to the north (beyond the mountains). In other parts of the coastal plain, however, the word for 'mountains' (*śḥɛr*) is used to indicate north instead. Another common way to describe south and north is to refer to the direction in which the water flows, with the result that the same word that means 'south' on the sea-side of the mountains can be used to indicate 'north' on the desert-side. However, all these rather complex sets of terms are being rapidly replaced, particularly in the speech of the younger generations and among urban populations, with the Arabic words for south and north (*ǧanūb* and *šimāl* respectively).

### 3.4.3 Color terms

The MSAL lexically encode four basic colour terms: white, black, red and green (**Bulakh2004**). For example, in Śḥehri one can find *lūn* for 'white', *ḥɔr* for 'black', *ʕɔ́ fər* for 'red' (and warm colours in general, including brown) and *śəẑrɔ́ r* for 'green' (and everything from green to blue). A fifth colour term, *ṣɔfrɔ̀ r* 'yellow' (Mehri *ṣāfər*), is most probably an adapted borrowing from Arabic already present at the common MSAL level (**Bulakh2004**).

A preliminary field inquiry on the subject was conducted by Gasparini in 2017, with 6 young speakers from the city of Salalah and its immediate surroundings,

all between 20 and 35 years old and all bilingual in Śḥehri and Arabic. The results of the tests showed a remarkable degree of idiolectal variation in the colour labeling systems employed by the informants, with different levels of interference from Arabic. Remarkably, when asked to label colours in Śḥehri from a printed basic color wheel which was shown to them during interviews, all the speakers used the Arabic word for 'blue', *azraq*, which seems to have replaced *śəẑrɔ́ r* (traditionally used for both blue and green, but now confined to the latter). Two speakers also used *axḍar* for 'green', claiming that they could not recall the Śḥehri term. In addition, only one speaker used *ʕɔ́ fər* for 'brown', Arabic *bunnī* being preferred by the other interviewees. The three basic colours 'white', 'black' and 'red', however, were regularly referred to using the Śḥehri forms by all speakers. Summing this up, it would seem that the Śḥehri colour system (at least in urban environments, but see below) is undergoing a radical restructuration. The three typologically "fundamental" colour terms are retained in most contexts, and a distinction between blue and green is being introduced through reduction of the original semantic spectrum of *śəẑrɔ́ r*, adoption of the Arabic word for blue, and subsequent replacement of *śəẑrɔ́ r* with *axḍar* (which indicates only green in Arabic). Further distinctions are either being replaced with the corresponding Arabic terms, or introduced if not part of the original semantic inventory of the language.

On this matter, Watson & Al-**Mahri2017** argue that color terms (together with numbers) are often among the first lexical items to be lost in contexts of linguistic endangerment, and that this is precisely the case with the MSAL. They write that even children in rural communities are now employing Arabic terms to refer to the different breeds of cattle (which traditionally used to be referred to by use of the three basic colour terms 'white', 'black' and 'red'). This is probably a result of the fact that even in villages younger generations are no longer involved in cattle herding. Examples include *aḥmar* 'bay' in place of Mehri *ōfar* or Śḥehri *ʕofer*, *aswad* 'black' in place of Mehri *ḥōwar*, and *abyaḍ* 'white' in place of Mehri *ūbōn*.

### 3.4.4 Other word classes

Watson & Al-**Mahri2017** note how, since the introduction of a public school system in Arabic in the 1970s, a number of common lexical items and expressions in Mehri and Śḥehri have been replaced by the corresponding Arabic ones. **Lonnet2009** also remarks how borrowings from Arabic are particularly common among particles and function words, Examples include *nafs aš-šī* 'the same thing' for Śḥehri *gens*, Mehri *gans*; *lākin* 'but' in place of Śḥehri *duʰn* and *min duʰn*,

Mehri *lahinnah*; *yaʕnī* 'that is to say' and *ʕabārah* in place of Śḥehri *yaxīn*, Mehri *(y)axah*; *tamām* 'fine' in place of Śḥehri *ḥayś̄ōf* and Mehri *hīs taww ~ histaww*; Mehri and Ḥarsusi *yā* 'oh... [voc.]' against MSAL *ʔā-*; Śḥehri *bdan*, Mehri *ʔàbdan* 'never, not at all', against Mehri and Ḥarsusi *bəhàwʔ*, Śḥehri *bhòʔ*. Consider also the case of Arabic *bəss* 'only', already mentioned in §**??** In Mehri as in Baṭhari, this particle appears now to be interchangeable with its equivalent *ār*, as example (**??**) shows:

(7) Mehri (**Sima2009**: 58:32, cited in **Watson2012**: 371; transcription adapted)

| bass | ta-ṭʕam=h | k̄ād axah / | ār |
|------|-----------|-------------|-----|
| only | 2s.m-to_taste.ipfv=3s.m | int / | only 2s.m-to_taste.ipfv |

ṭʕām ḏ-maḥḥ
of-clarified_butter

"Just taste it m. like it is just the taste of clarified butter"

As predictable, also in this field Baṭhari is the language most affected by Arabic: besides those already cited, we might add the expressions *zēn* 'well', *əbarr* 'outside', *xalaṣ* 'and this is it' (used to end a narrative). Finally, **Watson2012** remarks how 'Mahriyōt also exhibits structures unattested in Mehreyyet such as 'What X!' phrases reminiscent of Arabic, e.g. *maṭwalk* 'How tall you m.s. are!'".

## 4 Conclusions

Throughout this chapter we have repeatedly pointed out how research on the MSAL, and in particular on the effects that contact with Arabic has had on their evolution, is still far from reaching its mature stage. Much remains to be done, in particular, in terms of sheer documentation, especially in the case of the most endangered varieties (Hobyōt, Ḥarsūsi, Baṭhari). In addition to this, and although Watson's 2012 work has greatly contributed to expanding our knowledge in this area, MSAL syntax remains a strongly neglected field of inquiry. Finally, our knowledge of the history of the MSAL prior to the 20th century (and therefore the history of their contact with Arabic) is extremely poor.

It has also to be remarked that, although the most widely spoken among the MSAL are undoubtedly better documented, very little is known about the effects that urbanisation has had on their speech communities in recent years. In particular, anecdotal evidence suggests that the varieties of Śḥehri and Soqoṭri spoken in Salalah and Ḥadibo are undergoing rapid change under the influence of Arabic (both the standard variety of the language, which children learn in school,

and the dialects). Fieldwork conducted in the two abovementioned urban centres could provide extremely valuable information concerning the effects of contact between Arabic and Modern South Arabian.

Despite the far-from-complete state of the research in this field, what we currently know is sufficient to say that contact has had a strong impact on the MSAL. Though this is more evident in the area of lexicon, where borrowings are legion, phonetics and phonology have also been affected (though to a different extent from one language to another). Morphology and syntax, on the contrary, appear to be more resistant to contact-induced change, though in the most endangered varieties one can notice a partial disruption of the original pronominal system and verbal paradigm, and though the seemingly high degree of resistance to external influence shown by MSAL syntax could actually be due to our limited knowledge of the subject.

One last note is due concerning another heavily neglected topic, namely the effects that contact with the MSAL has had on spoken Arabic. Though we have not addressed the question in the course of this paper, evidence drawn from the existing literature (Simeone-**Senelle2002**) suggests that this influence, too, is not completely absent, and that further research in this direction could produce interesting results.

## Further reading

As we have seen, literature dealing specifically with the issue of contact between Arabic and the MSAL is extremely rare. **Morris2017** can be thought of as a general introduction to the topic. Watson and Al-**Mahri2017** offer an intriguing account of how language change, contact with Arabic and modification of the traditional environment are all deeply interrelated. Finally, **Lonnet2009** – although limited in scope and extension due to its nature as an encyclopedic entry – offers interesting highlights on the effects of contact on the MSAL.

```
%%move bib entries to  localbibliography.bib
```

```
@article{Beeston1981,
author = {Beeston, Alfred F. L.},
journal = {\textit{Arabica}},
number = {2/3},
pages = {178–186},
```

```
title = {Languages of Pre-{Islamic} Arabia},
volume = {28},
year = {1981}
}
```

Bulakh, Maria. 2004. Color Terms of the Modem South Arabian Languages: A Diachronic
282.

```
@book{Davey2016,
address = {Leiden},
author = {Davey, Richard J.},
number = {87},
publisher = {Brill},
series = {Studies in Semitic Languages and Linguistics volume},
title = {\textit{Coastal} \textit{Dhofari} \textit{{Arabic}:} \textit{a} \textit{s}
year = {2016}
}
```

```
@book{Dufour2016,
address = {\textit{Recherches} \textit{sur} \textit{le} \textit{verbe} \textit{sud}
author = {Dufour, Julien},
note = {(HDR - Habilitation à diriger des recherches [accreditation to supervise r
publisher = {École pratique des hautes études},
year = {2016}
}
```

```
@misc{Edzard2017,
author = {Edzard, Lutz},
note = {\textit{Brill's} \textit{Journal} \textit{of} \textit{Afroasiatic} \textit
2). 119–138.},
title = {On the role of Modern {South} Arabian within a comparative {Semitic} lexi
year = {2017}
}
```

```
@incollection{Gasparini2017,
address = {Turin},
author = {Gasparini, Fabio},
booktitle = {\textit{Linguistic} \textit{Studies} \textit{in} \textit{the} \text
Quaderni Di {RiCOGNIZIONI} 5)},
editor = {Bettega, Simone and Gasparini, Fabio},
pages = {69–85},
publisher = {Dipartmento di Lingue e Letterature straniere e Culture moderne, Un
title = {Phonetics of Emphatics in Bathari},
year = {2017a}
}


@book{Gasparini2017,
address = {\textit{The} \textit{Baṭḥari} \textit{Language} \textit{of} \textit{O
author = {Gasparini, Fabio},
note = {(Doctoral dissertation).},
publisher = {Università "L'Orientale" di Napoli},
year = {2017b}
}


@book{Hoyland2001,
address = {London},
author = {Hoyland, Robert G.},
publisher = {Routledge},
title = {\textit{Arabia} \textit{and} \textit{the} \textit{Arabs:} \textit{From}
year = {2001}
}


@article{Johnstone1975,
author = {Johnstone, Thomas M.},
journal = {\textit{Afro-Asiatic} \textit{Linguistics}}},
pages = {93–121},
title = {The modern {South} Arabian languages},
volume = {1},
year = {1975}
```

}


@book{Kogan2015,
address = {Berlin},
author = {Kogan, Leonid},
publisher = {Walter de Gruyter},
title = {\textit{Genealogical} \textit{Classification} \textit{of} \textit{{Semiti
year = {2015}
}


@book{LadefogedLadefoged1996,
address = {Oxford},
author = {Ladefoged, Peter and Maddieson, Ian},
publisher = {Blackwell},
title = {\textit{The} \textit{Sounds} \textit{of} \textit{the} \textit{World's} \te
year = {1996}
}


@incollection{Lonnet2009,
address = {Leiden},
author = {Lonnet, Antoine},
booktitle = {\textit{Encyclopedia} \textit{of} \textit{{Arabic}} \textit{Language}
editor = {Versteegh, Kees and Eid, Mushira},
pages = {296–300},
publisher = {Brill},
title = {{South} Arabian, Modern},
volume = {4},
year = {2009}
}


@misc{Lonnet2017,
author = {Lonnet, Antoine},
note = {\textit{Brill's} \textit{Journal} \textit{of} \textit{Afroasiatic} \textit
2). 265–290.},

*??*

title = {Modern {South} Arabian ikōtəb is not necessarily iparras or yənaggər},
year = {2017}
}


@incollection{Lucas2015,
address = {New York},
author = {Lucas, Christopher},
booktitle = {\textit{The} \textit{{{Routledge}}} \textit{handbook} \textit{of} \
editor = {Bowern, Claire and Evans, Bethwyn},
pages = {519–536},
publisher = {Routledge},
title = {Contact-Induced Language Change},
year = {2015}
}


@misc{Morris2017,
author = {Morris, Miranda J.},
note = {\textit{Brill's} \textit{Journal} \textit{of} \textit{Afroasiatic} \text
2). 9–32.},
title = {Some thoughts on studying the endangered Modern {South} Arabian Languag
year = {2017}
}


@book{NaumkinNaumkin1981,
address = {Moskow},
author = {Naumkin, Vitaly and Porkhomovsky, Victor},
publisher = {Nauka},
title = {\textit{Ocerki} \textit{po} \textit{etnolingvistike} \textit{sokotry} \
year = {1981}
}


@incollection{Persson2009,
address = {Wiesbaden},
author = {Persson, Maria},

booktitle = {\textit{Circumstantial} \textit{qualifiers} \textit{in} \textit{{Semi
editor = {Isaksson, Bo and Kammensjö, Heléne and Persson, Maria},
pages = {206--286},
publisher = {Harrassowitz},
title = {Circumstantial Qualifiers in Gulf {Arabic} Dialects},
year = {2009}
}


@article{Porkhomovsky1997,
author = {Porkhomovsky, Victor},
journal = {\textit{Proceedings} \textit{of} \textit{the} \textit{Seminar} \textit{
pages = {219–223},
title = {Modern {South} Arabian languages from a {{Semitic}} and Hamito-{{Semitic}}
volume = {27},
year = {1997}
}


@article{Retsö2000,
author = {Retsö, Jan},
journal = {\textit{Oriente} \textit{Moderno}},
number = {1},
pages = {111–118},
sortname = {Retso, Jan},
title = {"Kaškaša", T-Passives and the Ancient Dialects in Arabia},
volume = {19},
year = {2000}
}


@misc{RidouaneRidouane2017,
author = {Ridouane, Rachid and Gendrot, Cédric},
note = {\textit{Brill's} \textit{Journal} \textit{of} \textit{Afroasiatic} \textit
2). 139–159.},
title = {On ejective fricatives in {Omani} Mehri},
year = {2017}
}

```
@article{Robin1991,
author = {Robin, Christian-Julien},
journal = {\textit{Revue} \textit{des} \textit{mondes} \textit{musulmans} \texti
number = {1},
pages = {71–88},
title = {{La} pénétration des Arabes nomades au Yémen},
volume = {61},
year = {1991}
}


@book{Rubin2014,
address = {Leiden},
author = {Rubin, Aaron D.},
publisher = {Brill},
title = {\textit{The} \textit{Jibbali} \textit{(Shaḥri)} \textit{Language} \text
year = {2014}
}


@misc{Rubin2017,
author = {Rubin, Aaron D.},
title = {A Re-analysis of Mehri Grammar (Paper presented at the 5e journée d'étu
year = {2017}
}


@book{Sima2009,
address = {Wiesbaden},
author = {Sima, Alexander},
publisher = {Harrassowitz},
title = {\textit{A.} \textit{{Sima},} \textit{Mehri-Texte} \textit{aus} \textit{
year = {2009}
}
```

@incollection{Simeone-Senelle2002,
address = {Wiesbaden},
author = {Simeone-Senelle, Marie-Claude},
booktitle = {\textit{"Sprich} \textit{doch} \textit{mit} \textit{deinen} \textit{Kn
editor = {Jastrow, Otto and Arnold, Werner and Bobzin, Hartmut},
pages = {669--684},
publisher = {Harrassowitz},
title = {L'arabe parlé dans le Mahra (Yémen)},
year = {2002}
}


@incollection{Simeone-Senelle2011,
address = {Berlin},
author = {Simeone-Senelle, Marie-Claude},
booktitle = {\textit{The} \textit{{Semitic}} \textit{languages:} \textit{an} \texti
editor = {Weninger, Stefan and Khan, Geoffrey and Ungeheuer, Gerold and Wiegand, H
pages = {1073--1113},
publisher = {De Gruyter Mouton},
title = {Modern {South} Arabian},
year = {2011}
}


@book{Versteegh2014,
address = {Edinburgh},
author = {Versteegh, Kees},
publisher = {Edinburgh University Press},
title = {\textit{The} \textit{{Arabic}} \textit{Language}},
year = {2014}
}


@article{Watson2011,
author = {Watson, Janet C. E.},
journal = {\textit{Salford} \textit{Working} \textit{Papers} \textit{in} \textit{L
pages = {27--40},
title = {{South}~Arabian~and~Yemeni~dialects},

*??*

```
volume = {1},
year = {2011}
}


@book{Watson2012,
address = {Wiesbaden},
author = {Watson, Janet C. E.},
number = {52},
publisher = {Harrassowitz},
series = {Semitica Viva},
title = {\textit{The} \textit{structure} \textit{of} \textit{Mehri}},
year = {2012}
}


@incollection{WatsonWatson2017,
address = {Turin},
author = {Watson, Janet C. E. and Al-Mahri, Abdullah Musallam},
booktitle = {\textit{Linguistic} \textit{Studies} \textit{in} \textit{the} \text
Quaderni Di {RiCOGNIZIONI} 5)},
editor = {Bettega, Simone and Gasparini, Fabio},
pages = {87–103},
publisher = {Dipartimento di Lingue e Letterature straniere e Culture moderne, U
title = {Language and Nature in Dhofar},
year = {2017}
}


Watson, Janet C. E., Abdullah Musallam Al-Mahri \& Ali Al-Mahri. Forthcoming. \t

@article{WatsonWatson2010,
author = {Watson, Janet C. E. and Bellem, Alex},
journal = {\textit{Proceedings} \textit{of} \textit{the} \textit{Seminar} \texti
note = {. Oxford: Archaeopress.},
pages = {345–355},
title = {A detective story: {{{E}}}mphatics in Mehri},
volume = {40},
```

year = {2010}
}


@incollection{WatsonWatson2011,
address = {Amsterdam},
author = {Watson, Janet C. E. and Bellem, Alex},
booktitle = {\textit{Current} \textit{Issues} \textit{in} \textit{Linguistic} \text
editor = {Zeki, Majeed Hassan and Heselwood, Barry},
pages = {235–256},
publisher = {John Benjamins},
title = {Glottalisation and neutralisation in Yemeni {Arabic} and Mehri: {{A}}n ac
year = {2011}
}


@incollection{WatsonWatson2016,
address = {Amsterdam},
author = {Watson, Janet C. E. and Heselwood, Barry},
booktitle = {\textit{Studies} \textit{in} \textit{{Arabic}} \textit{Linguistics}},
editor = {Haddad, Youssef A. and Potsdam, Eric},
pages = {3–36},
publisher = {John Benjamins},
title = {Phonation and glottal states in Modern {South} Arabian and {San}'ani {Aral
year = {2016}
}


@article{WatsonEtAl2006,
author = {Watson, Janet C. E. and Glover Stalls, Bonnie and Al-razihi, Khalid and \
journal = {The language of Jabal Rāziḥ: Arabic or something else? \textit{Proceedir
pages = {35–41},
volume = {36},
year = {2006}
}


@article{Wellsted1837,

*??*

```
author = {Wellsted, James R.},
journal = {\textit{Journal} \textit{of} \textit{the} \textit{Royal} \textit{Geog
pages = {102--113},
title = {Narrative of a Journey into the Interior of Omán, in 1835},
volume = {7},
year = {1837}
}
```

# Chapter 17

# Berber

??

??

> Arabic has influenced Berber at all levels – not just lexically, but phonologically, morphologically, and syntactically – to an extent varying from region to region. Arabic influence is especially prominent in smaller northern and eastern varieties, but is substantial even in the largest varieties; only in Tuareg has Arabic influence remained relatively limited. This situation is the result of a long history of large-scale asymmetrical bilingualism often accompanied by language shift.

## 1 Current state and historical development

### 1.1 Introduction

Berber, or Tamazight, is the indigenous language family of northwestern Africa, distributed discontinuously across an area ranging from western Egypt to the Atlantic, and from the Mediterranean to the Sahel. Its range has been expanding within recent times in the Sahel as Tuareg speakers move southwards, but in the rest of this area, Berber has been present since before the classical period (Múrcia **Sánchez2010**). Its current discontinuous distribution is largely the result of language shift to Arabic over the past millennium.

At present, the largest concentrations of Berber speakers are found in the highlands of Morocco (Tashelhiyt, Tamazight, Tarifiyt) and northeastern Algeria (Kabyle, Chaoui). Tuareg, in the central Sahara and Sahel, is more diffusely spread over a large but relatively sparsely populated zone. Across the rest of this vast area, Berber varieties constitute small islands – in several cases, single towns – in a sea of Arabic.

This simplistic map, however, necessarily leaves out the effects of mobility – not limited to the traditional practice of nomadism in the Sahara and transhumance in parts of the Atlas mountains. The rapid urbanisation of North Africa

over the past century has brought large numbers of Berber speakers into traditionally Arabic-speaking towns, occasionally even changing the town's dominant language. The conquests of the early colonial period created small Berber-speaking refugee communities in the Levant and Chad, while more recent emigration has led to the emergence of urban Berber communities in western Europe and even Quebec.

## 1.2 Sociolinguistic situation of Berber

In North Africa proper, the key context for the maintenance of Berber is the village. Informal norms requiring the use of Berber with one's relatives and fellow villagers, or within the village council, encourage its maintenance not only there but in cities as well, depending on the strength of emigrants' (often multigenerational) ties to their hometowns. In some areas, such as Igli in Algeria (**Mouili2013**), the introduction of mass education in Arabic has disrupted these norms, encouraging parents to speak to their children in Arabic to improve their educational chances; in others, such as Siwa in Egypt (**Serreli2017**), it has had far less impact. Beyond the village, in wider rural contexts such as markets, communication is either in Berber or in Arabic, depending on the region; where it is in Arabic, it creates a strong incentive for bilingualism independent of the state's influence. For centuries, Berber-speaking villages in largely Arabic-speaking areas have sporadically been shifting to Arabic, as in the Blida region of Algeria (El **Arifi2014**); the opposite is also more rarely attested, as near Tizi-Ouzou in Algeria (**Gautier1913**).

In urban contexts, on the other hand, norms enforcing Berber have no public presence – quite the contrary. There one addresses a stranger in Arabic, or sometimes French, but rarely in Berber, except perhaps in a few Berber-majority cities such as Tizi-Ouzou (**Tigziri2008**). Even within the family, Arabic takes on increasing importance; in a study of Kabyle Berbers living in Oran (Algeria), Ait **Habbouche2013** found that 54% said they mostly spoke Arabic to their siblings, and 10% even with their grandparents. In the Sahel, Arabic is out of the picture, but there too family language choice is affected; 13% of the Berber speakers interviewed by **Jolivet2008** in Niamey (Niger) reported speaking no Tamasheq at all with their families, using Hausa or less frequently Zarma instead.

Bilingualism is widespread but strongly asymmetrical. Almost all Berber speakers learn dialectal Arabic (as well as standard Arabic, taught at school), whereas Arabic speakers almost never learn Berber. There are exceptions: in some contexts, Arabic-speaking women who marry Berber-speaking men need to learn Berber to speak with their in-laws (the author has witnessed several Kabyle ex-

amples), while Arabic speakers who settle in a strongly Berber-speaking town – and their children – sometimes end up learning Berber, as in Siwa (Egypt). Nevertheless, most Arabic speakers place little value on the language, and some openly denigrate it; in Bechar (Algeria), anyone expressing interest in Berber can expect frequently to hear the contemptuous saying *əš-šəlḥa ma-hu klam wə-d-dhən ma hu l-idam* 'Shilha (Berber) is no more speech than vegetable oil is animal fat.' To further complicate the situation, French remains an essential career skill (except in Libya and Egypt), since it is still the working language of many ministries and companies; in some middle-class families, it is the main home language spoken with children.

On paper, Berber (Tamazight) is now an official language of Morocco (since 2011) and Algeria (since 2016), while Tuareg (Tamasheq/Tamajeq) is a recognised national language of Mali and Niger. In practice, "official language" remains a misleading term. Official documents are rarely, if ever, provided in Berber, and there is no generalized right to communicate with the government in Berber. However, Berber is taught as a school subject in selected Algerian, Moroccan, and (since 2012 or so) Libyan schools, while some Malian and Nigerien ones even use it as a medium of education. It is also used in broadcast media, including some TV and radio channels. Both Morocco and Algeria have established language planning bodies to promote neologisms and encourage publishing, with a view towards standardisation. The latter poses difficult problems, given that each country includes major varieties which are not inherently mutually intelligible.

Berber varieties have been written since before the 2nd century BC (**Pichler2007**), – although the language of the earliest inscriptions is substantially different from modern Berber and decipherable only to a limited extent – and southern Morocco has left a substantial corpus of pre-colonial manuscripts (**Boogert1997**); many other examples could be cited from long before people such as **Mammeri1976** attempted to make Berber a printed language. Nevertheless, writing seems to have had very little impact on the development of Berber as yet. Awareness of the existence of a Berber writing system – Tifinagh – is widespread, and often a matter of pride. However, most Berber speakers have never studied Berber, and do not habitually read or write in it in any script – with the increasingly important exception of social media and text messages, typically in Latin or Arabic script depending on the region. Efforts to create a standard literary Berber language have not so far been successful enough to exert a unifying influence on its dispersed varieties. In the North African context, this is often understood as implying that Berber is not a language at all – "language" (Arabic *luɣa*) being popularly understood in the region as "standardized written language".

## 1.3 Demographic situation of Berber

No reliable recent estimate of the number of Berber speakers exists; relevant data is both scarce and hotly contested. The estimates brought together by Kossmann (2013:29–36; 2011:1) suggest a range of 30%–40% for Morocco, 20%–30% for Algeria, 8% for Niger, 7% for Mali, about 5% for Libya, and less than 1% for Tunisia, Egypt, and Mauritania. Selecting the midpoint of each range, and substituting in the mid-2017 populations of each of these countries (CIA 2017) would yield a total speaker population of about 25 million, 22 million of them divided almost evenly between Morocco and Algeria.

# 2 Contact languages

## 2.1 Across North Africa

Berber contact with Arabic began in the 7th century with the Islamic conquests. For several centuries, language shift seems to have been largely confined to major cities and their immediate surroundings, probably affecting Latin speakers more than Berber speakers. The invasion of the Banū Hilāl and Banū Sulaym in the mid-11th century is generally identified as the key turning point: it made Arabic a language of pastoralism, rapidly reshaping the linguistic landscape of Libya and southern Tunisia, then over the following centuries slowly transforming the High Plateau and the northern Sahara in general. This rural expansion further reinforced the role of Arabic as a lingua franca, while the recruitment of Arabic-speaking soldiers from pastoralist tribes encouraged its spread further west to the Moroccan Gharb.

The resulting linguistic divide between rural groups and towns remained a key theme of Maghrebi sociolinguistics until the 20th century. In several cases, a town spoke a different language than its hinterland; in much of the Sahara, Berber-speaking oasis towns such as Ouargla or Igli formed linguistic islands in regions otherwise populated by Arabic speakers, and in the North, towns such as Bejaia or Cherchell constituted small Arabic-speaking communities surrounded by a sea of Berber-speaking villages. Even in larger cities such as Algiers or Marrakech, the dominance of Arabic was counterbalanced by substantial regular immigration from Berber-speaking regions further afield.

Today all Berber communities are more or less multilingual, usually in Arabic and often also in French; outside of the most remote areas, monolingual speakers are quite difficult to find. Even in the 19th century, however, monolingual Berber speakers were considerably more numerous (**Kossmann2013**).

Alongside the coexistence of colloquial Maghrebi Arabic with Berber, Classical Arabic also had a role to play as the primary language of learning and in particular religious studies. Major Berber-speaking areas such as Kabylie (northern Algeria) and the Souss (southern Morocco) developed extensive systems of religious education, whose curricula consisted primarily of Arabic books (**Boogert1997**; **Mechehed2007**). The restriction of Classical Arabic to a limited range of contexts, and the relatively small proportion of the population pursuing higher education, gave it a comparatively small role in the contact situation; even in the lexicon, its influence is massively outweighed by that of colloquial Arabic, and it appears to have had no structural influence at all.

## 2.2 In Siwa

Examples of contact-induced change in this chapter are often drawn from Siwi, the Berber language of the oasis of Siwa in western Egypt. Sporadic long-distance contact with Arabic there presumably began in the seventh or eighth century with the Islamic conquests, and increased gradually as Lower Egypt and Cyrenaica became Arabic-speaking and as the trade routes linking Egypt to West Africa were re-established. During the eleventh century, the Banu Sulaym, speaking a Bedouin Arabic dialect, established themselves throughout Cyrenaica.

In the twelfth century, al-Idrīsī reports Arab settlement within Siwa itself, alongside the Berber population. Later geographers make no mention of an Arab community there, suggesting that these early immigrants were integrated into the Berber majority. Several core Arabic loans in Siwi, such as the negative copula *qačči* < *qaṭṭ šayʔ* and the noon prayer *luli* < *al-ʔūlē*, are totally absent from surrounding Arabic varieties today; such archaisms are likely to represent founder effects dating back to this period (**Souag2009**).

The available data gives nothing close to an adequate picture of the linguistic environment of medieval Siwa. We may assume that, throughout these centuries, most Siwis – or at least the dominant families – would have spoken Berber as their first language, and more mobile ones – especially traders – would have learned Arabic (but whose Arabic?) as a second language. Alongside these, however, we must envision a fluctuating population of Arabic-speaking immigrants and West African slaves learning Berber as a second language. In such a situation, both Berber-dominant and Arabic-dominant speakers should be expected to play a part in bringing Arabic influences into Siwi.

The oasis was integrated into the Egyptian state by Muhammad Ali in 1820, but large-scale state intervention in the linguistic environment of the oasis only took effect in the 20th century; the first government school was built in 1928,

and television was introduced in the 1980s. An equally important development during this period was the rise of labour migration, taking off in the 1960s as Siwi landowners recruited Upper Egyptian labourers, and Siwi young men found jobs in Libya's booming oil economy. It has then grown further since the 1980s with the rise of tourism and the growth of tertiary education. The effects of this integration into a national economy include a conspicuous generation gap in local second-language Arabic: older and less educated men speak a Bedouin-like dialect with *q > *g*, while younger and more educated ones speak a close approximation of Cairene Arabic.

# 3 Contact-induced changes in Berber

## 3.1 Introduction

As noted above, bilingualism in North Africa has been asymmetrical for many centuries, with Berbers much more likely to learn Arabic than vice versa. This suggests the plausible general assumption that the agents of contact-induced change were typically dominant in the (Berber) recipient language rather than in Arabic. However, closer examination of individual cases often reveals a less clear-cut situation; as seen above in §2.1, the history of Siwi suggests that Berber- and Arabic-dominant speakers both had a role to play, and *post facto* analysis of the language's structure seems to confirm this assumption. The loss of feminine plural agreement, for example (§3.3 below), can more easily be attributed to Arabic-dominant speakers adopting Berber than to Berber-dominant speakers. In the absence of clear documentary evidence, caution is therefore called for in the application of Van **Coetsem1988**'s (**Coetsem1988**) model to Berber.

## 3.2 Phonology

The influence of Arabic on Berber phonology is conspicuous; in general, every phoneme used in a given region's dialectal Arabic is found in nearby Berber varieties. Almost all northern Berber varieties have adopted from Arabic at least the pharyngeals /ʕ/ and /ḥ/, a series of voiceless emphatics: /ṣ/, /ḫ/, non-geminate /q/, and either /ḍ/ or /ṭ/. These phonemes presumably reached Berber through loanwords from Arabic, but have been extended to inherited vocabulary as well, through reinterpretation of emphatic spread or through their use in "expressive formations" (**Kossmann2013**), e.g. Kabyle *θi-ḥəðmər-θ* 'breast of a small animal' < *iðmar-ən* 'breast'.

In Siwi (**Souag2013**:36–39; Souag & van **Putten2016**), at least nine phonemes were clearly introduced from Arabic. The pharyngealised coronals /ṣ/, /ḷ/, /ṛ/ and /ḍ/ have no regular source in Berber, and occur in inherited vocabulary almost exclusively as a result of secondary emphasis spread (with the isolated exception of *ḍəs* 'to laugh'). The order of borrowing appears to be *ḷ, ṛ > ṣ > ḍ*; in a few older loans, Arabic *ṣ* is borrowed as *ẓ* (eg *ẓəffaṛ* 'to whistle' < *ṣaffar*), and in all but the most recent strata of loans, Arabic *ḍ/ð̣* is borrowed as *ṭ* (e.g. *a-ʕṛiṭ* 'broad' < *ʕarīḍ*). The pharyngeals /ḥ/ and /ʕ/ (e.g. *ḥəbba* 'a little' < *ḥabba* 'a grain', *ʕammi* 'paternal uncle' < *ʕamm-ī* 'uncle-OBL.1SG') likewise have no regular source in Berber, although 1SG -*ɣ*- has become -*ʕ*- for some speakers (an irregular sound change specific to this morpheme). *ʕ* is lost in a number of older loans (e.g. *an-naš* 'bier' < *an-naʕš*), but *ḥ* is always retained as such rather than being dropped or adapted (unlike Tuareg, where it is typically adapted to *ḫ*). This suggests that Siwi continued to adapt Arabic loans to its phonology by dropping *ʕ* up to some stage well after the beginning of significant borrowing from Arabic, but started accepting Arabic loans with *ḥ* too early for any adapted to survive, implying an order of borrowing *ḥ > ʕ*. Among the glottals, /h/ (e.g. *ddhan* 'oil' < *dihān* 'oils') appears in inherited vocabulary only in the distal demonstratives, where comparison to Berber languages that do have *h* suggests that it is excrescent, while /ʔ/ only rarely appears even in recent loanwords (e.g. *ʔəǧǧəṛ* 'to rent' < *ʔaǧǧar*). The mid vowel /o/ has been integrated into Siwi phonology as a result of borrowing from Arabic; having been established as a phoneme, however, it went on to emerge by irregular change from original *\*u* in two inherited words (*allon* 'window', *agṛoẓ* 'palm heart'), and from irregular simplification of *\*aɣu* in some demonstratives (eg *wok* 'this.SG.M' < *\*wa ɣuṛ-ək* 'this.sg.M at-2SG.M"). The interdentals /θ/ and /ð/ have a more marginal status, but are used by some speakers even in morphologically well-integrated loans, e.g. *a-θqil* or *a-tqil* 'heavy' < *θaqīl*. Arabic influence may also be responsible for the treatment of [ʒ] and [dʒ] as free variants of the same phoneme /ǧ/ (**Vycichl2005**), so that e.g. /taǧlaṣt/ 'spider' is variously realized as [tʰæʒlˤɑsˤt] ~ [tʰædʒlˤɑsˤt] (**Naumann2012**:152 SEE OTHER DOC); other Berber languages with phonemic *ž* normally have [dʒ] as a conditioned allophone (e.g. when geminated) or as a cluster.

Arabic influence has also massively affected the frequency of some phonemes. /q/ and /ḫ/ were marginal in Siwi before Arabic influence, while *\*e* had nearly disappeared due to regular sound changes, but all three are now quite frequent. Conversely, the influx of Arabic loans has helped make labiovelarised phonemes such as *gʷ*, *qʷ* rare.

### 3.3 Morphology

Berber offers numerous examples of the borrowing of Arabic words together with their original Arabic inflectional morphology, a case of what **Kossmann2010** calls Parallel System Borrowing. This phenomenon is most prominent for nominal number marking, but sometimes attested in other contexts too.

In Berber, most nouns are consistently preceded by a prefix marking gender (masculine/feminine), number (singular/plural), and often case/state. Nouns borrowed from Arabic normally either get assigned a Berber prefix, or fill the prefix slot with an invariant reflex of the Arabic definite article: cp. Figuig *a-gʕud* vs. Siwi *lə-gʕud* 'young camel' (< *qaʕūd*). The Berber plural marking system prior to Arabic influence was already rather complex, combining several different types of affixal marking with internal ablaut strategies; many Arabic loans are integrated into this system, e.g. Kabyle *a-bellar* 'crystal' > pl. *i-bellar-en* (< *billawr*), Siwi *a-kəddab* 'liar' > pl. *i-kəddab-ən* (< *kaððāb*). However, in most Berber varieties, Arabic loans have further complicated the system by frequently retaining their original plurals, e.g. Kabyle *l-kaɣeḍ* 'paper' > *le-kwaɣeḍ* (< *kāɣid*), Siwi *əl-gənfud* 'hedgehog' > pl. *lə-gnafid* (< *qunfuð*). (The difference correlates fairly well with the choice in the singular between a Berber prefix and an Arabic article, but not perfectly; contrast eg Siwi *a-fruḥ* 'chick, bastard' < *farḥ*, which takes the Arabic-style plural *lə-fraḥ*.) Berber has no inherited system of dual marking, instead using analytic strategies. Nevertheless, for a limited number of measure words, duals too are borrowed, e.g. Kabyle *yum-ayen* 'two days' < *yawm-ayn* (although 'day' remains *ass!*), Siwi *s-sən-t* 'year' > *sən-t-en* 'two years' < *san-at-ayn*. Arabic number morphology may sporadically spread to inherited terms as well, e.g. Kabyle *berdayen* 'twice' < *a-brid* 'road, time', Siwi *lə-gʷrazən* 'dogs' < *a-gʷərzni* 'dog' (**Souag2013**).

Whereas nouns are often borrowed together with their original inflectional morphology, verbs almost never are. The only attested exception is Ghomara, a heavily mixed variety of northern Morocco. In Ghomara, many (but not all) verbs borrowed from Arabic are systematically conjugated in Arabic in otherwise monolingual utterances, a phenomenon which seems to have remained stable over at least a century: thus 'I woke up' is consistently *faq-aḥ*, but 'I fished' is equally consistently *ṣṣað-iθ* (**Mourigh2016**:6, 137, 165). However, the borrowing of Arabic participles to express progressive aspect is also attested in Zuwara, if only for the two verbs of motion *mašəy* 'going' (pl. *mašy-iṇ*) and *žay* 'coming' (pl. *žayy-in*), contrasting with inherited *fəl* 'go', *asəd* 'come' (**Kossmann2013**:284–285).

Prepositions are less frequently borrowed; in some cases where they are bor-

rowed, however – including Igli *mənɣir-* 'except', Ghomara *bin* 'between' (**Kossmann2013**) – they too occasionally retain Arabic pronominal markers, e.g. Siwi *msabb-ha* 'for her' < *min sababi-hā* 'from reason.OBL-OBL.3SG.F' (**Souag2013**). In Awjili, more unusually, two inherited prepositions somewhat variably take Arabic pronominal markers, e.g. *dit-ha* 'in front of her' (Van **Putten2014**:113 SEE OTHER DOC).

A rarer but more spectacular example of morpheme borrowing is the borrowing of productive templates from Arabic. Such cases include the elative template əCCəC in Siwi, used to form the comparative degree of triliteral adjectives irrespective of etymology – thus *əmləl* 'whiter' < *a-məllal* alongside *əṭwəl* 'taller' < *a-ṭwil* < Arabic *ṭawīl* (**Souag2009**) – and the diminutive template CCiCəC in Ghomara (**Mourigh2016**), e.g. *aẓwiyyəṛ* 'little root' < *aẓaṛ* alongside *ləmwiyyəs* 'little knife' < *l-mus* < Arabic *al-mūsā* 'razor' (gemination of *y* is automatic in the environment *i_V*). As the latter example illustrates, borrowed derivational morphology sometimes becomes productive.

The effects of Arabic on Berber morphology are by no means limited to the borrowing of morphemes. There is reason to suspect Arabic influence of having played a role in processes of simplification attested mainly in peripheral varieties, such as the loss of case marking in many areas. In Siwi, where Arabic influence appears on independent grounds to be unusually high, the verbal system shows a number of apparent simplifications targeting categories absent in sedentary Arabic varieties: the loss of distinct negative stems, the near-complete merger of perfective with aorist, the fixed postverbal position of object clitics, and so on. It is tempting to explain such losses as arising from imperfect acquisition of Siwi by Arabic speakers.

Structural calquing in morphology is also sporadically attested. Siwi has lost distinct feminine plural agreement on verbs, pronouns, and demonstratives, extending the inherited masculine plural forms to cover plural agreement irrespective of gender. Within Berber, this is unprecedented; plural gender agreement is extremely well conserved across the family. However, it perfectly replicates the usual sedentary Arabic system found in Egypt and far beyond.

## 3.4 Syntax

Syntactic influence is often difficult to identify positively. Nevertheless, Berber offers a number of examples, of which relative clause formation is one of the clearest (**Souag2013**:151–156; **Kossmann2013**:369–407). Relative clauses in Berber are normally handled with a gap strategy combined with fronting of any stranded prepositions, as in (1).

(1) Awjila (**Paradisi1961**)

| ərrafəqa-nnəs | wi | ižin-an-a | nettin | id-sin | ksum |
|---|---|---|---|---|---|
| friend.PL-GEN.3SG | REL.PL.M | divide-3PL.M-PRF | 3SG.M | with-OBL.3PL.M | meat |

"his friends with whom he divided the meat"

In subject relativisation, a special form of the verb not agreeing in person (the so-called "participle") is used, as in (2); such a form is securely reconstructible for proto-Berber (**Kossmann2003**).

(2) Awjila (**Paradisi1960**)

| amədən | wa | tarəv-ən | nettin | ʕayyan |
|---|---|---|---|---|
| man | REL.SG.M | write.IPFV-PTCP | 3SG.M | ill |

"The man who is writing is ill."

In several smaller easterly varieties apart from Awjila, however, both of these traits have been lost. The strategy found in varieties such as Siwi – resumptive weak (affixal) pronouns throughout, and regular finite agreement for subject relativisation – perfectly parallels Arabic:

(3) Siwi
(**Souag2013**:151–152)

| tálti | tən | dəzz-ɣ-as | ǧǧəwab | "the woman to whom I sent |
|---|---|---|---|---|
| woman | REL.SG.F | send-1SG-DAT.3SG | letter | the letter" |

(4) (field data)

| aˊ | ggʷid | wənn | i-ʕəṃṃaṛ | iməǧran |
|---|---|---|---|---|
| man | REL.SG.M | 3SG.M-make.IPFV | sickle.PL | |

"the man who makes sickles"

In the case of verbal negation, an originally syntactic calque has often been morphologised in parallel in Arabic and Berber. A number of varieties – especially the widespread Zenati subgroup of Berber, ranging from eastern Morocco to northern Libya – have developed a postverbal negative clitic *š / ša* from *\*kăra* 'thing', apparently a calque on Arabic *š / ši* from *šayʔ*; however, some instead use the direct borrowings *ši* or *šay* (**Lucas2007**; **Kossmann2013**:332–334).

## 3.5 Lexicon

Lexical borrowing from Arabic is pervasive in Berber. Out of 41 languages around the world compared in the Loanword Typology Project (**Tadmor2009**), Tarifiyt

Berber was second only to (Selice) Romani in the percentage of loanwords – more than half (51.7%) of the concepts compared. More than 90% of loanwords examined in Tarifiyt were from Arabic, almost all from dialectal Maghrebi Arabic. There is little reason to suppose that Tarifiyt is exceptional in this respect among northern Berber languages; to the contrary, **Kossmann2013** finds its rate of basic vocabulary borrowing to be typical of northern Berber, whereas Siwi and Ghomara go much higher. The rate of borrowing from Arabic, however, is considerably lower further south and west; on a 200-word list of basic vocabulary, Chaker (1984:225–226) finds 38% Arabic loans in Kabyle (north-central Algeria) vs. 25% in Tashelhiyt (southern Morocco) and only 5% in Tahaggart Tuareg (southern Algeria).

This borrowing is pervasive across the languages concerned, rather than being restricted to particular domains. Every semantic field examined for Tarifiyt, including body parts, contained at least 20% loanwords, and verbs or adjectives were about as frequently borrowed as nouns were (**Kossmann2009**). Numerals stand out for particularly massive borrowing; most northern Berber varieties have borrowed all numerals from Arabic above a number ranging from 1 to 3 (**Souag2007**).

The effects of this borrowing on the structure of the lexicon remain insufficiently investigated, but appear prominent in such domains as kinship terminology. Throughout northern Berber, a basic distinction between paternal kin and maternal kin is expressed primarily with Arabic loanwords (*ʕammi* 'paternal uncle' vs. *xali* 'maternal uncle' etc.), whereas in Tuareg that distinction is not strongly lexicalised. Nevertheless, borrowing does not automatically entail lexical restructuring; Tashelhiyt, for example, kept its vigesimal system even after borrowing the Arabic word for 'twenty' (*ʕšrin*), cf. Ameur (2008:77 SEE OTHER DOC).

The borrowing of analysable multi-word phrases – above all, numerals followed by nouns – stands out as a rather common outcome of Berber contact with Arabic. Usually this is limited to the borrowing of numerals in combination with a limited set of measure words, such as 'day'; thus in Siwi we find forms like *sbaʕ-t iyyam* 'seven days' rather than the expected regular formation *\*səbʕa n nnḥaṛ-at (**Souag2013**).* In Beni Snous (western Algeria), the phenomenon seems to have gone rather further: **Destaing1907** reports that numerals above 10 systematically select for Arabic nouns. **SouagKherbache2016**, however, explain this as a code-switching effect rather than a true case of one language's grammar requiring shifts into another.

# 4 Conclusion

The influence of Arabic on Berber has come to be better understood over the past couple of decades, but much remains to be done. Synchronically, Berber–Arabic codeswitching remains virtually unresearched; rare exceptions include **Hamza2007** and **Kossmann2012**. Sociolinguistic methods could help us better understand the gradual integration of new Arabic loanwords; the early efforts of **Brahmi2000** have hardly been followed up on. Diachronically, it remains necessary to move beyond the mere identification of loanwords and contact effects towards a chronological ordering of different strata, an approach explored for some peripheral varieties by **Souag2009** and van **PuttenBenkato2017**. While linguists are belatedly beginning to take advantage of earlier manuscript data to understand the history of Berber (**Boogert1997**; **Boogert1998**; **Brugnatelli2011**; **Meouak2015**), this data has not yet been used in any systematic way to help date the effects of contact at different periods. For many smaller varieties, especially in the Sahara, basic documentation and description are still necessary before the influence of Arabic can be explored. The unprecedented degree of Arabic influence revealed in Ghomara by recent work (**Mourigh2016**), extending to the borrowing of full verb paradigms, suggests that such descriptive work may yet yield dividends in the study of contact.

Despite all these gaps, the work done so far is more than sufficient to establish a general picture of Arabic influence on Berber. Throughout northern Berber, Arabic influence on the lexicon is substantial and pervasive, bringing with it significant effects on phonology and morphology. Structural effects of Arabic on morphology, and Arabic influence on Berber syntax, are less conspicuous but nevertheless important, especially in smaller varieties such as Siwi. Looking at these results through Van Coetsem's framework, this suggests that RL-dominant speakers have had an especially dominant role in Arabic–Berber contact in larger varieties, whereas SL-dominant speakers' role is more visible in smaller varieties. However, this *a priori* conclusion should be tested against directly attested historical data wherever possible.

## Further reading

The key reference for Arabic influence on northern Berber is **Kossmann2012**, frequently cited above; this covers all levels of influence including the lexicon, phonology, nominal and verbal morphology, borrowing of morphological categories, and syntax. The most extensive in-depth study of Arabic influence on

a specific Berber variety is **Souag2013**, effectively a contact-focused grammatical sketch of Siwi Berber. **Mourigh2016** is a thorough synchronic description of by far the most strongly Arabic-influenced Berber variety, Ghomara, giving a uniquely clear picture of just how far the process can go without resulting in language shift.

## Acknowledgements

## Abbreviations

DAT dative
  F feminine
  GEN genitive
  M masculine
  IPFV imperfective
  PRF perfect
  PL plural
  PTCP participle
  SG singular

```
%%move bib entries to  localbibliography.bib
\end{stylelsUnNumberedSection}


\begin{styleBibliography}
@book{AitHabbouche2013,
address = {Oran},
author = {Ait Habbouche, Khadidja},
note = {http://theses.univ-oran1.dz/document/TH3963.pdf.},
publisher = {University of Oran Mémoire de Magister},
title = {Language Maintenance and Language Shift among Kabyle Speakers  in {Arabic}
year = {2013}
}
```

*??*

```
\begin{styleBibliography}
@book{Boogert1997,
address = {Leiden},
author = {Boogert, Nico van den},
note = {}},
publisher = {Nederlands Institu\textsc{ut} \textsc{voor} \textsc{het} \textsc{Na
title = {\textit{The} \textit{Berber} \textit{Literary} \textit{Tradition} \text
year = {1997}
}
```

```
\begin{styleBibliography}
@book{Boogert1998,
address = {Aix-en-Provence},
author = {Boogert, Nico van den},
publisher = {Institut de recherches et d'études sur le monde arabe et musulman (
title = {\textit{"{La}} \textit{révélation} \textit{des} \textit{énigmes" :} \te
year = {1998}
}
```

```
\begin{styleBibliography}
@incollection{Brahimi2000,
address = {Berlin},
author = {Brahimi, Fadila},
booktitle = {\textit{{Arabic}} \textit{as} \textit{a} \textit{Minority} \textit{
editor = {Jonathan Owens},
pages = {371–382},
publisher = {Mouton de Gruyter},
title = {Loanwords in Algerian Berber},
year = {2000}
}
```

```
\begin{styleBibliography}
```

```
@incollection{Brugnatelli2011,
address = {Milano},
author = {Brugnatelli, Vermondo},
booktitle = {\textit{Studies} \textit{on} \textit{Language} \textit{and} \textit{{A
editor = {Luca Busetto},
note = {A.S.A.R. http://unimib.academia.edu/VermondoBrugnatelli/Papers/1304580/Some
pages = {29--40},
publisher = {Qu},
title = {Some grammatical features of Ancient {Eastern} Berber (the language of the
year = {2011}
}
```

```
\begin{styleBibliography}
@book{Chaker1984,
address = {Paris},
author = {Chaker, Salem},
publisher = {Editions du Centre national de la recherche scientifique},
title = {\textit{Textes} \textit{en} \textit{linguistique} \textit{berbère :} \text
year = {1984}
}
```

```
\begin{styleBibliography}
@book{CIA2017,
address = {Washington, D.C.},
author = {CIA},
publisher = {Central Intelligence Agency},
title = {\textit{The} \textit{World} \textit{\citealt{Factbook2017}}},
url = {https://www.cia.gov/library/publications/the-world-factbook/index.html},
year = {2017}
}
```

```
\begin{styleBibliography}
```

*??*

```
@book{Coetsem1988,
address = {(Publications in Language Sciences 27). Dordrecht, Holland ; Providen
author = {Coetsem, Frans van},
publisher = {Foris Publications},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two}
year = {1988}
}
```

```
\begin{styleBibliography}
@book{Destaing1907,
address = {Paris},
author = {Destaing, Edmond},
publisher = {Ernest Leroux},
title = {\textit{Etude} \textit{sur} \textit{le} \textit{dialecte} \textit{berbè
year = {1907}
}
```

```
\begin{styleBibliography}
@book{ElArifi2014,
address = {4e ed. Web},
author = {El Arifi, Samir},
note = {http://atlas.blida.over-blog.com/lexique.tachelhit.tamazight.atlas.blide
publisher = {Haut Commissariat à l'Amazighité},
title = {\textit{Tachelhit} \textit{de} \textit{l'Atlas} \textit{blidéen}},
year = {2014}
}
```

```
\begin{styleBibliography}
@article{Gautier1913,
author = {Gautier, E.-F},
journal = {\textit{Annales} \textit{de} \textit{géographie}},
note = {. doi:10.3406/geo.1913.8304.},
```

```
number = {123},
pages = {255–266},
title = {Répartition de la langue berbère en Algérie},
volume = {22},
year = {1913}
}




\begin{styleBibliography}
@book{Hamza2007,
address = {Brighton},
author = {Hamza, Belgacem},
note = {http://ethos.bl.uk/OrderDetails.do?did=1\& uin=uk.bl.ethos.444351.},
publisher = {University of Sussex PhD thesis},
title = {Berber ethnicity and language shift in {Tunisia}},
year = {2007}
}




\begin{styleBibliography}
@incollection{Jolivet2008,
address = {Köln},
author = {Jolivet, Remi},
booktitle = {\textit{Berber} \textit{in} \textit{Contact:} \textit{Linguistic} \te
editor = {Mena Lafkioui and Vermondo Brugnatelli},
pages = {139–149},
publisher = {Rüdiger Köppe},
title = {{La} fidélité au tamashek des berbérophones dans le contexte multilingue
year = {2008}
}




\begin{styleBibliography}
@book{Kossmann2003,
address = {München},
```

*??*

author = {Kossmann, Maarten},
publisher = {LINCOM Europa},
title = {The origin of the Berber "participle." \textit{Selected} \textit{Compar
40},
year = {2003}
}


\begin{styleBibliography}
@incollection{Kossmann2009,
address = {Berlin},
author = {Kossmann, Maarten},
booktitle = {\textit{Loanwords} \textit{in} \textit{the} \textit{World's} \texti
editor = {Martin Haspelmath and Uri Tadmor},
pages = {191–214},
publisher = {de Gruyter},
title = {Loanwords in Tarifiyt},
year = {2009}
}


\begin{styleBibliography}
@article{Kossmann2010,
author = {Kossmann, Maarten},
journal = {\textit{Diachronica}},
note = {. doi:10.1075/dia.27.3.03kos.},
number = {3},
pages = {459–487},
title = {Parallel System Borrowing: {{P}}arallel morphological systems due to th
volume = {27},
year = {2010}
}


\begin{styleBibliography}

```
@book{Kossmann2011,
address = {1., Aufl. (Berber Studies 30). Köln},
author = {Kossmann, Maarten},
publisher = {Rüdiger Köppe},
title = {\textit{A} \textit{Grammatical} \textit{Sketch} \textit{of} \textit{Ayer}
year = {2011}
}



\begin{styleBibliography}
Kossmann, Maarten. 2012. Berber-Arabic code-switching in Imouzzar du Kandar. (Ed.)
382.


\begin{styleBibliography}
@book{Kossmann2013,
address = {Leiden},
author = {Kossmann, Maarten},
publisher = {Brill},
title = {\textit{The} \textit{{Arabic}} \textit{Influence} \textit{on} \textit{{No
year = {2013}
}



\begin{styleBibliography}
@article{Lucas2007,
author = {Lucas, Christopher},
journal = {\textit{Transactions} \textit{of} \textit{the} \textit{Philological} \t
note = {. doi:10.1111/j.1467-968X.2007.00189.x.},
number = {3},
pages = {398–431},
title = {{Jespersen}'s cycle in {Arabic} and Berber},
volume = {105},
year = {2007}
}
```

*??*

\begin{styleBibliography}
@book{Mammeri1976,
address = {Paris},
author = {Mammeri, Mouloud},
publisher = {Maspéro},
title = {\textit{Tajerrumt} \textit{n} \textit{tmazight} \textit{(tantala} \text
year = {1976}
}


\begin{styleBibliography}
@book{Mechehed2007,
address = {L'organisation des notices de catalogage des manuscrits arabes ou ber
author = {Mechehed, Djamel-Eddine},
note = {\textit{Les} \textit{manuscrits} \textit{berbères} \textit{au} \textit{M
publisher = {cas de la collection Ulahbib, Kabylie},
year = {2007}
}


\begin{styleBibliography}
@book{Meouak2015,
address = {Brill. doi},
author = {Meouak, Mohamed},
note = {1163/9789. http://booksandjournals.brillonline.com/content/books/9789 (5
publisher = {10},
title = {\textit{{La}} \textit{langue} \textit{berbère} \textit{au} \textit{Magh
urldate = {004302358},
year = {2015}
}


\begin{styleBibliography}

```
@article{Mouili2013,
author = {Mouili, Fatiha},
journal = {\textit{Humanities} \textit{and} \textit{Social} \textit{Sciences} \tex
number = {2},
pages = {563--580},
title = {The Berber language of Igli, language towards extinction},
volume = {2},
year = {2013}
}




\begin{styleBibliography}
@book{Mourigh2016,
address = {(Berber Studies 45). Köln},
author = {Mourigh, Khalid},
publisher = {Rüdiger Köppe},
title = {\textit{A} \textit{Grammar} \textit{of} \textit{Ghomara} \textit{Berber} \
year = {2016}
}




\begin{styleBibliography}
@book{MúrciaSánchez2010,
address = {Barcelona},
author = {Múrcia Sánchez, Carles},
note = {http://hdl.handle.net/.},
publisher = {Universitat de Barcelona},
sortname = {Murcia Sanchez, Carles},
title = {{La} llengua amaziga a l'antiguitat a partir de les fonts gregues i llati
urldate = {10803/1724},
year = {2010}
}




\begin{styleBibliography}
```

*??*

```
@misc{Paradisi1960,
author = {Paradisi, Umberto},
note = {\textit{Rivista} \textit{degli} \textit{Studi} \textit{Orientali} XXXV.
177.},
title = {Il berbero di Augila, Materiale lessicale},
year = {1960}
}




\begin{styleBibliography}
@misc{Paradisi1961,
author = {Paradisi, Umberto},
note = {\textit{Annali} \textit{dell'Istituto} \textit{Universitario} \textit{Or
91.},
title = {Testi berberi di Augila (Cirenaica)},
year = {1961}
}




\begin{styleBibliography}
@book{Pichler2007,
address = {Köln},
author = {Pichler, Werner},
publisher = {Rüdiger Köppe},
title = {\textit{Origin} \textit{and} \textit{Development} \textit{of} \textit{t
year = {2007}
}




\begin{styleBibliography}
@incollection{PuttenPutten2017,
address = {Leiden},
author = {Putten, Marijn van and Adam Benkato},
booktitle = {\textit{{{Arabic}}} \textit{in} \textit{Context:} \textit{Celebrati
editor = {Ahmad Al-Jallad},
```

```
pages = {476–502},
publisher = {Brill},
title = {The {Arabic} Strata in Awjila Berber},
year = {2017}
}




\begin{styleBibliography}
@incollection{Serreli2017,
address = {Abingdon},
author = {Serreli, Valentina},
booktitle = {\textit{Identity} \textit{and} \textit{Dialect} \textit{Performance:}
editor = {Reem Bassiouney},
pages = {226–242},
publisher = {Routledge},
title = {Language and identity in Siwa Oasis: {{I}}ndexing belonging, localness an
year = {2017}
}




\begin{styleBibliography}
@incollection{Souag2007,
address = {Cambridge},
author = {Souag, Lameen},
booktitle = {\textit{Proceedings} \textit{for} \textit{the} \textit{Fifth} \textit
editor = {Naomi Hilton, Rachel Arscott, Katherine Barden, Sheena Shah and Meg Zell
pages = {237–244},
publisher = {Cambridge Institute of Language Research},
title = {The Typology of Number Borrowing in Berber},
url = {http://www.ling.cam.ac.uk/camling/Manuscripts/CamLing2007\_Souag.pdf},
year = {2007}
}




\begin{styleBibliography}
```

*??*

```
@article{Souag2009,
author = {Souag, Lameen},
journal = {\textit{Zeitschrift} \textit{für} \textit{Arabische} \textit{Linguist
pages = {51–75},
title = {Siwa and its significance for {Arabic} dialectology},
volume = {51},
year = {2009}
}
```

```
\begin{styleBibliography}
@book{Souag2013,
address = {(Berber Studies 37). Köln},
author = {Souag, Lameen},
publisher = {Rüdiger Köppe},
title = {\textit{Berber} \textit{and} \textit{{Arabic}} \textit{in} \textit{Siwa
year = {2013}
}
```

```
\begin{styleBibliography}
@article{SouagSouag2016,
author = {Souag, Lameen and Fatma Kherbache},
journal = {\textit{International} \textit{Journal} \textit{of} \textit{Bilingual
note = {. doi:10.006914536002.},
number = {2},
pages = {97–115},
title = {Syntactically conditioned code-switching? {{T}}he syntax of numerals in
urldate = {1177/1367},
volume = {20},
year = {2016}
}
```

```
\begin{styleBibliography}
```

```
@misc{SouagSouag2016,
author = {Souag, Lameen and Marijn van Putten},
note = {\textit{Studies} \textit{in} \textit{African} \textit{Linguistics} 45(1\& 2
208.},
title = {The origin of mid vowels in {Siwi}},
year = {2016}
}
```

```
\begin{styleBibliography}
@incollection{Tadmor2009,
address = {Berlin},
author = {Tadmor, Uri},
booktitle = {\textit{Loanwords} \textit{in} \textit{the} \textit{World's} \textit{
editor = {Martin Haspelmath and Uri Tadmor},
pages = {55–75},
publisher = {De Gruyter},
title = {Loanwords in the world's languages: {{F}}indings and results},
year = {2009}
}
```

```
\begin{styleBibliography}
Tigziri, Nora. 2008. Le kabyle au contact des langues en présence en Algérie : Entr
185. Köln: Rüdiger Köppe.
```

```
\begin{styleBibliography}
@book{Vycichl2005,
address = {\textit{Berberstudien ;} \textit{\&} , \textit{A} \textit{Sketch} \text
147. Köln},
author = {Vycichl, Werner},
publisher = {Rüdiger Köppe},
title = {Sprachskizzen: {{D}}ie berberischen Sprachinseln in Tunesien},
year = {2005}
}
```

*??*

# Chapter 18

# Beja–Arabic contact

??

??

> This chapter argues for two types of outcomes of the long-standing and intense contact situation between Beja and Arabic in Sudan: borrowings at the phonological, syntactic and lexical levels, convergence at the morphological level.

## 1 Current state and historical development

### 1.1 Historical development of Beja

Beja is the sole language of the northern Cushitic branch of the Afroasiatic phylum. Recent archeological discoveries show growing evidence that Beja is related to the extinct languages of the Medjay (from which the ethnonym 'Beja' is derived (Rilly, 2014:1175), and Blemmye tribes, first attested on Egyptian inscriptions of the 12th Dynasty for the former, and on a Napatan stela of the late 7th cent. BCE for the latter; for recent discussions, see **Browne2003**; El-**Sayed2011**; Zibelius-**Chen2014**; **Rilly2014**, Rilly (forth. [2018]). The Medjays were nomads living in the eastern Nubian Desert, between the 1st and 2nd cataracts of the Nile River. The Blemmyes invaded and took part in defeating the Meroitic kingdom, fought against the Romans up to the Sinai, and ruled Nubia from Talmis (modern Kalabsha, between Luxor and Aswan) for a few decades, before being defeated themselves by the Noubades around 450 CE (Rilly, forth. [2018]). In Late antiquity, the linguistic situation involved, in northern Lower Nubia, Cushitic languages, north-eastern Sudanic languages, to which Meroitic and Nubian belong, also Coptic and Greek to some extent, and in the south, Ethio-Semitic. It is likely that there was mutual influence to an extent that is difficult to disentangle today.

## 1.2 Current situation of Beja

The Beja territory has shrunk a lot since Late antiquity, and Beja (*biḍawije:=t*) is mainly spoken today in the Red Sea and Kassala States in eastern Sudan, in the dry lands between the Red Sea and the Atbara River. **The1993** census, the last one to include a language question, recorded some 1,100,000 Beja speakers, and there is probably at least double that figure today. There are also some 60,000 speakers in northern Eritrea and there may be still a few speakers left in Egypt, in the Nile valley at Aswan and Daraw, and along the coast towards Marsa ʿAlam (Morin, 1995; Wedekind, 2012). In Sudan today, Beja speakers have also settled in Khartum and cities in central and western Sudan (Hamid Ahmed, 2005a:67).

All Bejas today are Muslims. They consider themselves as Bedouins, and call themselves *arab* 'Arab'[1]; they call the ethnical Arabs *balawje:=t*. Before the introduction of modern means of transportation, they were traditionally the holders of the caravan trade in the desert towards the west, south and north of their territory, and they still move between summer and winter pastures with their cattle. They also produce sorghum and millet for daily consumption, and fruits and vegetables in the oases. The arrival of Rashaida migrants from Saudi Arabia in the 19th century created tensions in an area with meagre resources, but the first contemporary important social changes took place during the British mandate with the agricultural development of the Gash and Tokar areas, and the settlement of non-Beja farmers. The draughts of the mid-1980s brought about a massive exodus towards the cities, notably Port-Sudan and Kassala, followed by job diversification, and increased access to education in Arabic, although not generalized, especially for girls who rarely go beyond primary level (Hamid Ahmed, 2005a).

Beja is mostly an oral language. In Eritrea, a Latin script has been introduced in schools after independence in 1993, but in Sudan no education in Beja exists. The attempts made by the SIL and at the University of the Red Sea to implement an Arabic-based script remained with no future. On the other hand, since the last few years, school teachers in rural areas, tend to talk more and more in Beja in order to fight illiteracy (in Arabic) and absenteeism (Onour, 2015).

---

[1]In Sudan the term *ʕarab* is widely used for referring to nomad groups in general, and not only to ethnically defined Arabs. Thanks to Stefano Manfredi for this information.

## 2 Contact languages

Contact between Bejas and Arabs started as early as the beginning of Islamization and through the trade relations with Muslim Egypt and the Arab penetration in search of gold and emerald. Evidence of these contacts lie in the early arabicization of Beja anthroponyms (Záhořík, 2007). The date of the beginning of Islamization differs according to authors, but it seems it started as early as the 10th century, and slowly expanded until it became the sole religion between the 18th and the 19th centuries (Záhořík, 2007).

We have no information concerning the beginning and spreading of Beja-Arabic bilingualism. It is thus often impossible to figure out if a transfer occurred through Beja-dominant speakers or was imposed by fluent Beja-Arabic bilingual speakers, and consequently to tell apart whether a contact-induced feature belongs to the borrowing or to the imposition type of transfer as advocating by Van **Coetsem1988** and his followers. What is sure though, is that socio-historical as well as linguistic evidence speak in favour of Beja-Arabic bilingualism as an ancient phenomena, but in unknown proportion among the population. With the spread of Islam since the Middle Ages, contact with Arabic became more and more prevalent in Sudan.

In this country, which will be the focus of this chapter, bilingualism with Sudanese Arabic is frequent, particularly for men, and expanding, including among women in cities and villages but to a lesser extent. Bejas in Port-Sudan are also in contact with varieties of Yemeni Arabic. Rural Bejas recently settled at the periphery of the big cities have the reputation of being more monolingual than others, which was still the case fifteen years ago (Vanhove, 2003).

The Beja language is an integral part of the social and cultural identity of the people, but it is not a necessary component. Tribes and clans that have switched to Tigre, such as the Beni Amer, or Arabic are considered Bejas. Beja is prestigious since it permits to stick to the ethical values of the society, and is considered as aesthetically pleasing due to its allusiveness character. The attitude towards Arabic is ambivalent. It is perceived as taboo-less, and thus contrary to the rules of honour, nevertheless its use allows to avoid transgressing them. Arabic is also prestigious because it is the language of social promotion and modernity (Hamid Ahmed, 2005b). The language attitudes are rapidly changing, and there is some concern among the Beja diaspora about the future of the Beja language, even though it cannot be considered as endangered. Some parents avoid speaking Beja to their children, in fear that it would interfere with their learning of Arabic at school, leaving to the grand-parents the transmission of Beja (Wedekind, 2012;

Vanhove, 2017). But there is no reliable quantitative or qualitative sociolinguistic study of this phenomenon. Codeswitching between Beja and Arabic is spreading but understudied.

This sketchy description of the sociolinguistic situation of Beja speaks for at least two types of transfer: (i) borrowing, where the agents of transfer are dominant in the recipient language (Beja); (ii) convergence phenomena, since the difference in linguistic dominance between the languages of the bilingual speakers tends to be really small (at least among male speakers today, and probably earlier in the history of Beja) (see van **Coetsem1988**:87). Imposition has probably also occurred of course, but it is not always easy to prove.

## 3 Contact-induced changes in Beja

### 3.1 Phonology

The few contact-induced changes in Beja phonology belong to the borrowing type.

The phonological system of Beja counts 21 consonantal phonemes, presented in Table 1.

Table 1: Beja consonants.

|   | f | t | ʈ |   | k | kʷ | ʔ |
|---|---|---|---|---|---|---|---|
| b |   | d | ɖ |   | g | gʷ |   |
|   |   | s |   | ʃ |   |   | h |
|   |   |   |   | ʤ |   |   |   |
| m |   | n |   |   |   |   |   |
|   |   | l |   |   |   |   |   |
|   |   | r |   |   |   |   |   |
| w |   |   |   | j |   |   |   |

The voiced post-alveolar affricate *ʤ* (often realised as a voiced palatal plosive *[ɟ]* as in Sudanese Arabic) deserves attention as a possible outcome of contact with Arabic. Since (Reinisch, 1893:17), it is usually believed that this affricate is only present in Arabic loanwords and is not a phoneme (Roper, 1928; Hudson, 1976; Morin, 1995). The existence of a number of minimal pairs in word-initial position invalidates the latter analysis: *ʤi:k* 'rooster' ~ *ʃi:k* 'chewing tobacco'; *ʤhar* 'chance' ~ *dhar* 'bless'; *ʤaw* 'quarrel' ~ *ɖaw* 'jungle' ~ *ʃaw* 'preg-

nancy' ~ *gaw* 'house' (Vanhove, 2017). As for the former claim, there are actually a few lexical items such as *bʔaʤi* 'bed', *gʷʔaʤi* 'one-eyed' (*gʷʔad* 'two eyes'), that cannot be traced back to Arabic (the second one is pan-Cushitic, Blažek, ms). Nevertheless, it is the case that most items containing this phoneme do come from (or through) (Sudanese) Arabic: *a:laʤ* 'tease', *aʤi:n* 'dough', *aʤib* 'please', *ʔaʤala* 'bicycle', *ʔiʤir* 'divine reward', *ʤa:hil* 'small child', *ʤabana* 'coffee', *ʤalla:j* 'because of', *ʤallab* 'fish', *ʤanna* 'paradise', *ʤanta:ji* 'djinn', *ʤarika:n* 'jerrycan', *ʤe:b* 'pocket', *ʤhali:* 'coal', *ʤimʔa* 'week', *ʤins* 'sort', *ʤuwwa* 'inside', *faʤil* 'morning', *finʤa:n* 'cup', *hanʤar* 'dagger', *hiʤ* 'pilgrimage', *maʤaʔa* 'famine', *maʤlis* 'reconciliation meeting', *siʤin* 'prison', *tarʤima:l* 'translator', *waʤʤa* 'appointment', *xawaʤa* 'foreigner'. It is clear that *ʤ* is not marginal anymore. However *ʤ* is unstable: it has several dialectal variants, *tʃ*, *g* and *d*, and may alternate with the dental *d* or retroflex *ɖ*, in the original Beja lexicon (*ʤiwʔo:r / ɖiwʔo:r* 'honorable man') as well as in loanwords (*aʤi:n / aɖi:n* 'dough') (Vanhove and Hamid Ahmed, 2011; Vanhove, 2017). In my data, which counts some 50 male and female speakers of all age groups, this is rarely the case, meaning that there is a good chance that this originally marginal phoneme will live on under the influence of (Sudanese) Arabic.

There are two other consonants in Arabic loanwords that are regularly used by the Beja speakers: *z* and *x*, neither of which can be considered as phonemes since there are no minimal pairs.

Blažek (2007a:130) established a regular correspondence between Beja *d* and Proto-East-Cushitic *\*z*. In contemporary Beja *z* only occurs in recent loanwords from Sudanese Arabic such as *ʤaza* 'wage', *ʤo:z* 'pair', *rizg* 'job', *wazʔ* 'offer', *xazna* 'treasure', *zama:n* 'time', *zirʔa* 'field', *zu:r* 'visit'. It may alternate with *d*, even within the speech of the same speaker as free variants, e.g. *dama:n, dirʔa, du:r*. The fricative alveolar pronunciation is more frequent among city dwellers, which are more often bilingual. It is difficult to ascertain whether Beja is in the process of re-acquiring the voiced fricative through contact with Sudanese Arabic, or whether it will undergo the same evolution to a dental stop as in the past.

A few recent Arabic loanwords may also retain the voiceless velar fricative *x* (see also **ManfrediEtAl2015**:304-305): *xazna* 'treasure', *xawaʤa* 'foreigner', *xadda:m* 'servant', *xa:tar* 'be dangerous', *a:xar* 'last'. In my data, this is usually the case in the speech of fluent bilingual speakers. We have thus here a probable imposition type of transfer. In older borrowings, even among these speakers, Arabic *x* shifted to *h* (*xajma > he:ma* 'tent'). It may be because these older loans spread in a community which was at that time composed mainly of Beja-dominant speakers, but we have no means of proving this hypothesis.

## 3.2 Morphology

### 3.2.1 General remarks

Most Cushitic languages only have concatenative morphology, the stem and pattern schema being at best highly marginal (Cohen, 1988:256). In addition to Beja, Afar and Saho (Lowland East-Cushitic (LEC) branch), Beja's geographically closest sisters, are exceptions, and all three languages use also non-concatenative morphology. In Afar and Saho it is by far less pervasive than in Beja, in particular they do not use vocalic alternation for verbal derivation and this feature is restricted to the verb flexion for a minority of underived verbs.

Even though Beja and Arabic share a similar type of morphology, the following overview shows that each language has developed its own system. Although they have been in contact for centuries, neither small scale nor massive borrowing from Arabic morphological patterns can be postulated for the Beja data. An interpretation in terms of a convergence phenomenon is more relevant, both in terms of semantics and forms.

Non-concatenative morphology concerns an important portion of the lexicon: a large part of the verb morphology (conjugations, verb derivations, verb-noun derivations), and part of the noun morphology (adjectives, nouns, "internal" plurals, and to a lesser extent, place and instrument names). In what follows, I build on **Vanhove2012**, and **Vanhove2017**, correcting some inaccuracies.

### 3.2.2 Verb morphology

Only one of the two Beja verb classes, the one conjugated with prefixes (or infixes), belongs to non-concatenative morphology. This verb class (V1) is formed of a stem which undergoes ablaut varying with TAM, person and number, to which prefixed personal indices for all TAMs are added (plural and gender morphemes are also suffixes). V1 is diachronically the oldest pattern, which survives only in a few other Cushitic languages. In Beja V1s are the majority (57%), as against approximately 30% in Afar and Saho, and only five verbs in Somali and South Agaw (Cohen, 1988:256). Table 2 provides examples in the Perfective and Imperfective for bi-consonantal and tri-consonantal roots.

Prefix conjugations are used in Arabic varieties and South-Semitic languages but their functions and origins are different. In South-Semitic, prefix conjugation has an aspectual value of Imperfective while in Cushitic it marks a particular morphological verb class. The Cushitic prefix conjugation (in the singular) goes back to auxiliary verbs meaning 'say' or 'be', while the prefix conjugation of

Table 2: Perfective and Imperfective patterns

|  | | **Bi-consonantal *dif* 'go'** | **Tri-consonantal *kitim* 'arrive'** |
|---|---|---|---|
| PFV | *i-dif* 'he went' | *i-ktim* 'he arrived' | |
| | *i-dif-na* 'they went' | *i-ktim-na* 'they arrived' | |
| IPFV | *i-n-di:f* 'he goes' | *k\<an\>ti:m* 'he arrives' | |
| | *e:-dif-na* 'they go' | *e:-katim-na* 'they arrive' | |

South-Semitic has various origins, none of them including a verb 'say' or 'be' (Cohen, 1984). Although different grammaticalization chains took place in the two branches of Afroasiatic, this suggests that the root and pattern system might have already been robust in Beja at an ancient stage of the language. It is noteworthy that there are at least traces of vocalic alternation between the Perfective and the Imperfective in all Cushitic branches (Cohen, 1984:88-102), thus reinforcing the hypothesis of an ancient root and pattern schema in Beja. In what proportion this schema was entrenched in the morphology of proto-Cushitic lexicon is impossible to decide.

Verb derivation of V1s is also largely non-concatenative. Beja is the only Cushitic language which uses qualitative ablaut in the stem for the formation of semantic and voice derivation. The ablaut can combine with prefixes.

Table 3 presents the five verb derivation patterns with ablaut, and Table 6 shows the absence of correspondence between the Beja and Arabic (classical and Sudanese) patterns. Sudanese patterns are extracted from **Bergman2002**, who does not provide semantic values.

Among the Semitic languages, an intensive pattern similar to the Beja one is only known in the modern South Arabian (MSA) languages spoken in Eastern Yemen (not in contact with Beja), where it is also used for causation and transitivisation. The MSA languages are direct cognates of Ethio-Semitic and it is usually admitted that they were imported by South-Arabian speakers (Ullendorf, 1955). Still this ablaut pattern was not retained in Ethio-Semitic. It is also unknown in Cushitic. In classical Arabic, the plurisyllabic pattern does not have an intensive value, but a goal or sometimes reciprocal meaning.

Beja is the sole Cushitic language which differentiates between active and mid-

Table 3: V1 derivation patterns with ablaut

|  | **Monosyllabic V1** | **Plurisyllabic V1** |
|---|---|---|
| INT | *boːs* (*bis* 'burry') | *kaːtim* (*kitim* 'arrive') |
| MID | *faf* (*fif* 'pour') | *rimad* (*rimid* 'avenge') |
| PASS | *aːtoː-maːn* (*min* 'shave') | *at-dabaːl* (*dibil* 'gather') |
|  |  | *am-heːjid* (*haːjid* 'sew') |
| RECP | *amoː-gaːd* (*gid* 'throw') | *am-garaːm* (*girim* 'be in-imical') |
| CAUS |  | *si-katim* (*kitim* 'arrive') |

dle voices by means of vocalic alternation. Remnants of this pattern exist in some Semitic languages, among them Arabic, in a fossilized form.

In Cushitic, qualitative ablaut for the passive voice only occurs in Beja. Passive formation through ablaut exists in classical and Sudanese Arabic, but with different vowels. **Bergman2002** mentions that "a handful of verbs in S[udanese] A[rabic]" can be formed that way. For Stefano Manfredi (p.c.) it is a productive pattern in this Arabic variety.

Like the passive voice, the reciprocal is characterized by a qualitative ablaut in *aː* in the stem, but the prefix is different and consists of *am(oː)-*. *m* is not used for verbal derivation in Arabic, which uses the same ablaut, but for the first vowel of disyllabic stems, to express, marginally, the reciprocal of the base form. Most often the reciprocal meaning is expressed by other forms with the *t-* prefixed or infixed to the derived form or the base form. In some other Cushitic languages *-m* is used as a suffix for passive or middle voice (without ablaut). In Beja *m-* can also marginally be used as a passive marker, together with ablaut, for a few transitive intensive verbs: *ameː-saj* 'be flayed', *ameː-biḍan* 'be forgotten'.

Although a suffix *-s* (not a prefix as in Beja) is common in Cushitic, Beja is once more the only Cushitic language which uses ablaut for the causative derived form. Neither ablaut nor the *s-* prefix exist in Arabic. Arabic uses different patterns for the causative: the same as the intensive one, i.e. with a geminated second root consonant, and the (ʔ)a-CcaC(a) pattern.

This brief overview shows that Beja has not borrowed patterns from (Sudanese) Arabic, but has at best similar, but not exact, cognate patterns which are marginal in both classical and Sudanese Arabic.

Beja has also four non-finite verb forms. The simultaneity converb of V1s is the only one with non-concatenative morphology. The affirmative converb is

Table 4: Comparison between Beja and Arabic derivation patterns

| | Beja Pluri-syllabic V1 | Classical Arabic | Sudanese Arabic |
|---|---|---|---|
| INT | Ca:Ca *ka:tim* < *kitim* 'arrive' | CaCCaCa, Ca:-CaCa | CaCCaC, Ca:CaC |
| MID | CiCaC *rimad* < *rimid* 'avenge' | CuCiCa, iC<t>aCaCa, ta-CaCCaCa, ta-Ca:CaCa, in-CaCaCa, tas-CaCaCa, ista-CaCaCa | it-CaCCaC, it-CaCaC |
| PASS | at-CaCa:C *at-daba:l* < *dibil* 'gather' am-Ce:CiC *am-he:jid* < *ha:jid* 'sew' | CuCiCa, iC<t>aCaCa, ta-CaCCaCa, ta-Ca:CaCa, in-CaCaCa, ista-CaCaCa | it-Ca:CaC, it-CaCaC, CiCiC, in-CaCaC |
| RECP | am-CaCa:C *am-gara:m* < *girim* 'be inimical' | Ca:CaCa , ta-Ca:CaCa, iC<t>aCaCa | Ca:CaC, it-Ca:CaC |
| CAUS | si-CaCiC *si-katim* < *kitim* 'arrive' | CaCCaCa, ʔa-CCaCa | CaCCaC, a-CCaC |

marked for both verb classes with a suffix *-e:* added to the stem: *gid* 'throw', *gid-e:* 'while throwing'; *kitim* 'arrive', *kitim-e:* 'while arriving'. In the negative polarity, the negative particle *ba:=*, precedes the stem, and V1s undergo ablaut in the stem (Ci:C and CaCi:C), and drop the suffix; it has a privative meaning: *ba:=gi:d* 'without throwing'; *ba:=kati:m* 'without arriving'. No similar patterns exist in Arabic or other Cushitic languages.

### 3.2.3 Verb-noun derivation

The non-concatenative morphology concerns only V1s. It applies to action nouns (masdars) and agent nouns.

There are several masdar patterns, with or without a prefix, with or without ablaut, depending mostly on the syllabic structure of the verb. The most frequent ones with ablaut are presented below.

The pattern *m(i(:)/a)*-CV(:)C applies to the majority of monosyllabic verbs. The stem vowel varies and is not predictable: *di* 'say', *mi-ja:d* 'saying'; *dir* 'kill', *ma-dar* 'killing'; *sʔa* 'sit down', *ma-sʔa:* 'sitting'; *ak* 'become', *mi:-kti* 'becoming'; *hiw* 'give', *mi-jaw* 'gift, act of giving'. A few disyllabic V1s comply to this pattern: *rikʷij* 'fear', *mi-rkʷa:j* 'fearing'; *jiwid* 'curl', *mi:-wad* 'curling'. Some V1s of the CiC pattern have a Ca:C pattern for masdars, without a prefix: *gid* 'throw', *ga:d* 'throwing'. In classical Arabic, the marginal masdars with a prefix concern trisyllabic verbs, none showing a long vowel in the stem or the prefix, nor a vowel *i* in the prefix.

CiCiC and HaCiC[2] disyllabic verbs form their masdars by vocalic ablaut to *u:*: *kitim* 'arrive', *kitu:m* 'arriving'; *ʔabik* 'take', *ʔabu:k* 'taking'; *hamir* 'be poor', *hamu:r* 'being poor'. CiCaC V1s, and those ending in *-j* undergo vocalic ablaut to *e:*: *digʷagʷ* 'catch up', *digʷe:gʷ* 'catching up'; *biḍa:j* 'yawn', *biḍe:j* 'yawning'. In classical Arabic, the masdar pattern with *u:* has a different vowel in the first syllable, *a* (in Beja *a* is conditioned by the initial laryngeal consonant), and it is limited almost exclusively to verbs expressing movements and body positions (Blachère and Gaudefroy-Demombynes, 1975 (3rd ed.):81).

**Bergman2002** provides no information about verbal nouns of the base form in Sudanese Arabic except that they "are not predictable".

As for agent nouns of V1s, they most often combine ablaut with the suffix *-a:na*, the same suffix as the one used to form agent nouns of V2 verbs, whose stems do not undergo ablaut: *bir* 'snatch', *bo:r-a:na* 'snatcher'; *gid* 'throw', *ge:d-a:na* 'thrower, a good shot'; *dibil* 'pick up', *da:bl-ana* 'one who picks up'. Some tri-

---

[2]Where H stand for the laryngeals *ʔ* and *h*.

consonantal stems have a suffix *-i* instead of *-a:na*: *ʃibib* 'look at', *ʃa:bb-i* 'guard, sentinel. Some have both suffixes: *kitim* 'arrive', *ka:tm-a:na* / *ka:tim-i* 'newcomer'.

These patterns are unknown in Arabic.

### 3.2.4 Noun morphology

#### 3.2.4.1 General remarks

The existence of verb-noun derivation patterns and nominal plural patterns are well recognized in the literature about Beja morphology; for a recent overview, see **Appleyard2007**. It is far from being the case for adjective and noun patterns. All noun and adjective patterns linked to V1s are listed below. **Vanhove2012** provides an overview of these patterns which are summed up below.

#### 3.2.4.2 Adjective patterns

There are eight adjective patterns, two of which are shared with nouns. Most are derived from V1 verbs, but the reverse is also attested. A corresponding verb form is inexistent in a few cases. All patterns are based on ablaut, in two cases with an additional suffix *-a*, or gemination of the medial consonant. Arabic has no dedicated adjective pattern (but the active participle pattern of the verbal base form Ca:CiC may express properties). Table 5 provides the full list of patterns with examples. It is remarkable that none of them is similar to those of classical Arabic or colloquial Sudanese Arabic (**Bergman2002**).

Table 5: Adjective patterns

| Pattern | Adjective | Verb form |
|---------|-----------|-----------|
| aCa:C | *ama:g* 'bad' | *mig* 'do evil' |
| CaCCa | *marʔa* 'wide' | *mirʔ* 'be wide' |
| Ca:Ci(C) | *na:kʷis* 'short' | *nikʷis* 'be short' |
| | *da:ji* 'good' | Ø |
| CaCi(C) | *dawil* 'close' | *diwil* 'be close' |
| CaCa:C3 | *takʷa:kʷ* 'prepared' | *tikʷikʷ* 'prepare' |
| CaCa:C-a | *raga:g-a* 'long' | *rigig* 'stand up' |
| CiCa:C | *ʃikʷa:n* 'aromatic' | *ʃikʷan* 'emit pleasant odour' |
| CaCCiC | *ʃallik* 'few' | *ʃilik* 'be few' |

### 3.2.4.3 Nouns

There are eleven 'basic' noun patterns related to V1 verbs. Most of the patterns of triconsonantal roots resemble those of Arabic (but are not strictly identical), a coincidence which is not surprising since both languages have a limited number of vowels. Table **??** provides the full list of these patterns. The CaCi and CiCi patterns are shared with adjectives. The CiCi(C) pattern does not undergo ablaut.

Table 6: Noun patterns

| Pattern | Noun | Verb form |
|---------|------|-----------|
| CaC | *nak$^w$* 'pregnancy' | *nik$^w$i* 'become pregnant' |
| CiCa | *nisa* 'advise' | *nisa* 'advice' |
| CaCi | *sari* 'wakefulness' | *sir* 'keep awake' |
| CaCa | *nada* 'dew' | *nidaj* 'sweat, exude water' |
| CiCi(C) | *mirʔi* 'width' | *mirʔ* 'be wide' |
|  | *riʃid* 'wealth' | *riʃid* 'raise, care for cultivation or cattle' |
| CaCi:C | *ʃaḍi:ḍ* 'strip' | *ʃiḍiḍ* 'strip off' |
|  | *ʃak$^w$i:n* 'fragrance' | *ʃik$^w$an* 'emit pleasant odour' |
| CaCi:C-a | *raʃi:d-a* 'cattle' | *riʃid* 'raise, care for cultivation or cattle' |
| Ci:Ca:C | *ti:la:l* 'stride' | *tilil* 'stride far away from home' |
| CaCo:C | *tabo:k* 'double handful' | *tibo:k* 'fill scoop with two hands joined' |
| CiCu:C-a | *tilu:l-a* 'exile' | *tilil* 'stride far away from home' |
| CaCo:C | *tabo:k* 'double-handful' | *tibo:k* 'fill scoop with cupped hands' |
| CiCu:C-a | *tilu:l-a* 'exile' | *tilil* 'stride far away from home' |

### 3.2.4.4 Nouns with prefix *m(V)-*

A few other semantic types of nouns, mostly instrument and place names, are formed through ablaut and a prefix *m(V)-*, like in Arabic. Contrary to Arabic where these patterns are productive, they are frozen forms in Beja (some are not loanwords from Arabic, see the last three examples): *ʔafi* 'prevent, secure', *m-ʔafaj* 'nail, rivet, fastener'; *ginif* 'kneel', *mi-gnaf* 'camp'; *himi* 'cover', *m-himme:j* 'blanket'; *mo:k* 'take shelter', *ma-kʷa* 'shelter'; *rifif* 'drag an object on ground', *mi-rfaf* 'reptile'.

### 3.2.5 Plural patterns

The so-called "internal plural" patterns are common and frequent in Arabic (and Ethio-Semitic). Beja also has a limited set of internal plural patterns, but it has developed its own system. Ablaut patterns for plural formation mainly concern non-derived nouns containing either a long vowel or ending in a diphthong. Both *i:* and *u:* turn to *i* in the plural, and *a:*, *e:* and *o:* turn to *a*, sometimes with the addition of the plural suffix *-a*; nouns ending in *-aj* turn to a long vowel *-e:j*: *angʷi:l*, pl. *angʷil* 'ear'; *lu:l*, pl. *lil* 'rope'; *asu:l*, pl. *asil* 'blister'; *hasa:l*, pl. *hasal* / *hasal-a* 'bridle'; *me:k*, pl. *mak* 'donkey'; *bo:k*, pl. *bak* 'he-goat'; *ganaj*, pl. *gane:j* 'gazelle' (**Vanhove2017**).

Even though internal plurals can be considered as a genetic feature, the fact that they are very rare or absent in other Cushitic languages (**Zaborski1986**) speaks for a possible influence of Arabic (in Sudan) upon Beja.

## 3.3 Syntax

### 3.3.1 General remarks

As far as we know, there are no syntactic calques from Arabic in Beja. There are nevertheless a few lexical and grammatical borrowed items that gave rise to constructions concerning coordination and subordination.

### 3.3.2 Coordination

One of the three devices that mark coordination is borrowed from Arabic *wa*. It is only used for noun phrases or nominalized clauses (deranked, temporal and relative clauses), whereas the Arabic source particle can be used with noun phrases and simple sentences. *wa* is preposed to the coordinated element in Arabic, but in Beja it is an enclitic particle *=wa*, a position in line with the favoured SOV

word order. *=wa* follows each of the coordinated elements. (??) illustrates the coordination of two noun phrases.

(1) Beja (BEJ_MV_NARR_01_shelter_057)[3]
    bʔaɖaɖ=wa    i=koːlej=wa        sallam-ja=ajt=heːb
    sword=COORD DEF.M=stick=COORD give-PFV.3SG.M=CSL=OBJ.1SG
    "Since he had given me a sword and the stick…"

Deranked clauses with non-finite verb forms, which partly have nominal properties (Vanhove, 2016), are also coordinated with *=wa*. (??) is an example with the general converb, and (??) with the simultaneity converb.

(2) Beja (BEJ_MV_NARR_14_sijadok_281-284)
    winneːt si-raːkʷ-oːm-aː=b=wa
    plenty   CAUS-fear.INT-PASS-CVB.MNR=INDF.M.ACC=COORD
    gadab-aː=b=wa                    ʔas-ti          far-iːni
    be_sad-CVB.MNR=INDF.M.ACC=COORD be_up-CVB.GNRL jump-IPFV.3SG.M
    "Very frightened and sad, he jumps up."

(3) Beja (BEJ_MV_NARR_13_grave_126-130)
    afirh-a=b                      aka-jeː=wa
    be_happy-CVB.MNR=INDF.M.ACC become-CVB.SMLT=COORD
    i=dheːj=iːb              hawaː-jeː=wa          rh-ani
    DEF.M=people=LOC.SG play-CVB.SMLT=COORD see-PFV.1SG
    "I saw him happy and playing among the people."

Relative and temporal subordinate clauses also have nominal properties: the relative markers derive from the articles, and the temporal markers go back to nouns. (??) illustrates the coordination with a relative clause which bears the coordination marker, and (??) the coordination of two temporal clauses.

(4) Beja (06_foreigner_22-24)
    t=ʔarabijaːj=wa    oː=maːl                w=haːj
    DEF.F=car=COORD DEF.SG.M.ACC=treasure DEF.SG.M/REL=COM
    a-kati=eːb=wa              kass=oː
    come-CVB.MNR=INDF.M.ACC 1SG-become.IPFV=REL.M=COORD

---

[3]The sources of the examples are accessible online at http://corpafroas.huma-num.fr/Archives/ and http://cortypo.huma-num.fr/; the indications in parenthesis refer to the texts they are extracted from.

jʔ-a=b                    a-ni:w=ho:k
all=POSS.3SG.ACC 1SG-give.IPFV=OBJ.2SG

"I'll give you a car and all the fortune that I brought."

(5) Beja (BEJ_MV_CONV_01_rich_SP2_136-138)
na:=t          bi=i-hi:w=o:=ho:b=wa
thing=INDF.F NEG.OPT=3SG.M-give.OPT=OBJ.1SG=when=COORD
ba=a-di=e:k          i-ni:w=o:=ho:b=wa
NEG.OPT=1SG-say=if 3SG.M-give.IPFV=OBJ.1SG=when=COORD

"Whether he gives it to me or not..." (lit. when he does not give me anything and when he gives me)

Adversative coordination between two simple clauses is also expressed with a borrowing from Arabic: *la:kin* 'but'.

### 3.3.3 Subordination

The reason conjunction *sabbi* 'because' is a borrowing from the Arabic noun *sabab* 'reason'. Like most balanced adverbial clauses, it is based on one of the relative clause types, the one nominalized with the noun *na* 'thing' in the genitive case. *sabbi* functions as the head of the relative clause.

(6) Beja (03_camel_192)
ʔakir-a              ḍab  ḍa:b-i:n=e:=na:-ji              sabbi
be_strong-CVB.MNR run.AC run-AOR.3PL=REL=thing-GEN because

"Because it was running so fast..."

*sabbi* can also be used after a noun or a pronoun in the genitive case: *ombarijo:k sabbi* 'because of you'.

Terminative adverbial clauses are expressed with a borrowing from Arabic, *hadi:d* 'limit'. Again the borrowing is the head of the relative clause.

(7) Beja (BEJ_MV_NARR_51_camel_stallion_026-030)
o:n            i=ka:m=o:k                    he:=he:b
PROX.SG.M.ACC DEF.M=camel=POSS.2SG.ACC give.[IMP.SG.M]=OBJ.1SG
i-ndi          e:n baru:k        o:=bu:n
3SG.M-say.IPFV DM 2SG.M.NOM DEF.SG.M.ACC=coffee
gʷʔa-ti=e:b              hadi:d
drink-AOR.2SG.M=REL.M until

"Leave your camel with me, he says, until you have drunk your coffee!"

*hadi:d* can also be used as a postposition after a noun, in which case it can be abbreviated to *had*: *faʤil=had* 'until morning'.

## 3.4 Lexicon

The study of the Beja lexicon lacks research on the adaptation of Arabic loan-words and their chronological layers. There is no statistic on the proportions of lexical items borrowed from Arabic or Ethio-Semitic as compared to those inherited from Cushitic, not to mention Afroasiatic as a whole or borrowed from Nilo-Saharan. Phonetic and morphological changes are bound to have blurred the etymological data, but what is sure is that massive lexical borrowings from Arabic for all word categories took place at different periods of time, and that the process is still going on. Lexicostatistics studies (Cohen, 1988:267; Blažek, 1997) have shown that Beja shares only 20% of the basic vocabulary with its closest relatives, Afar, Saho and Agaw.

In this section I mainly concentrate on verbs, because they are often believed to be less easily borrowed in language contact situations (see **Wohlgemuth2009** for an overview of the literature on this topic), which obviously is not the case for Beja.

**Cohen1988** mentions that tri-consonantal V1s contain a majority of Semitic borrowings. I did a search in **Reinisch1895** dictionary, the only one to mention possible correspondences with Semitic languages. It provided a total of 225 V1s, out of which only nine have no Semitic cognates (four are cognates with Cushitic, one is borrowed from Nubian, and one cognate with Egyptian). Even if some of Reinisch's comparisons are dubious, the overall picture is still in favour of massive borrowings from Semitic (96%). It is not easy to disentangle whether the source is an Ethio-Semitic language or Arabic, but until a more detailed study can be undertaken, the following can be said: 55 verbs (20%) have cognates only in Ethio-Semitic (Tigre, Tigrinya, Amharic, and/or Ge'ez); out of the remaining 161 (72%), 85 are attested only in Arabic, 76 also in Ethio-Semitic. Because of the long-standing contact with Arabic for a large majority of Beja speakers in Sudan, and the marginality of contact with Tigre limited to the South of the Beja domain, it is tempting to assume that almost 3/4 of the 76 verbs are of Arabic origin. They may have been borrowed at an unknown time when the new suffix conjugation was still marginal. However, there are also tri-consonantal verbs (V2) which are conjugated with suffixes, albeit less numerous: 164. 141 have cognates with Semitic languages (95 Arabic, 31 Ethio-Semitic, and 15 attested in both branches), six are pan-Cushitic, one is pan-Afroasiatic, one Nubian, six are of dubious origin, and nine occur only in Beja. Does this mean that these borrowings

occurred later than for V1s? In the current state of our knowledge of the historical development of Beja, it is not possible to answer this question.

On the other hand, **Cohen1988**, in his count of consonants per stem in eight Cushitic and Omotic languages, showed that biconsonantal stems are predominant in six of the languages. On the contrary, they form 52.8% of the Beja 770 stems of **Roper1928** lexicon, and 42.7% of the Agaw 611 stems, almost on a par with bi-consonantal stems (42.2%). What this shows is that massive borrowings from Arabic (or from Ethio-Semitic for Agaw) helped to the preservation of triconsonantal stems which still form a majority of the stems in Beja as opposed to other Cushitic languages.

## 4 Conclusion

This overview has shown that massive lexical borrowings from Arabic in Beja has helped to significantly entrench non-concatenative morphology in this language. Whether this is a preservation of an old Cushitic system or a more important development of this structure than in other Cushitic languages under the influence of Arabic is open to debate, but what is certain is that it is not incidental that this system is so pervasive in Beja, the only Cushitic language to have had a long history of intense language contact with Arabic, the Semitic language where non-concatenative morphology is the most developed. What is important to recall is that Beja non-concatenative morphology shows no borrowings of Arabic patterns, leading to the conclusion that we are dealing with a convergence phenomenon. Lexical borrowings and morphological convergence are not paralleled in the phonological and syntactic domains where the Arabic influence seems marginal.

A lot remains to be done about language contact between Beja and Arabic and we lack reliable sociolinguistic studies in this domain. We also lack a comprehensive historical investigation of the Beja lexicon as well as a sufficiently elaborated theory of phonetic correspondences for Cushitic (Cohen, 1988:267). Even though important progress have been made, in particular for Beja in the comparison of its consonant system with other Cushitic languages and for the etymology of lexical items for some semantic fields, thanks to Blažek (ms., 2003a, 2003b, 2006, 2007b), the absence of a theory of lexical borrowings in Beja (and other Cushitic languages) is still an impediment for a major breakthrough in the understanding of language contact between Beja and Arabic.

# Further reading

Apart from **Vanhove2012** on non-concatenative morphology already summed up in §**??**, we lack studies on contact-induced changes in Beja. **Vanhove2003** is a brief article on code-switching in one tale and two jokes based on conversational analysis. **Wedekind2012** is an appraisal of the changing sociolinguistic situation of Beja in Egypt, Sudan and Eritrea.

# Acknowledgements

# Abbreviations

AC action noun, masdar; ACC accusative; AOR aorist; CAUS causative; COM comitative; COORD coordination; CSL causal; CVB converb; DEF definite; DM discourse MARKER; F feminine; GEN genitive; GNRL general; IMP imperative; INDF indefinite; INT intensive; IPFV imperfective; LOC locative; M masculine; MID middle; MNR manner; NEG negation; NOM nominative; OBJ object; OPT optative; PASS passive; PFV perfective; PL plural; POSS possessive; PROX proximal; RECP reciprocal; REL relator; SG singular; SMLT simultaneity.

```
%%move bib entries to  localbibliography.bib

@book{Appleyard2007,
address = {In \textit{Morphologies} \textit{of} \textit{Asia} \textit{and} \text
author = {Appleyard, David},
publisher = {Eisenbrauns},
title = {Beja morphology},
year = {2007}
}
```

```
@book{Bergman2002,
address = {Springfield, VA},
author = {Bergman, Elizabeth},
publisher = {Dunwoody Press},
title = {\textit{Spoken} \textit{Sudanese} \textit{{Arabic}:} \textit{Grammar,} \t
year = {2002}
}



@book{BlachèreGaudefroy-Demombynes1975,
address = {\textit{Grammaire} \textit{de} \textit{l'arabe} \textit{classique}. Par
author = {Blachère, Régis, and Gaudefroy-Demombynes, Maurice},
publisher = {Maisonneuve et Larose},
sortname = {Blachere, Regis, and Gaudefroy-Demombynes, Maurice},
title = {(3e ed.)},
year = {1975}
}



@book{Blažek1997,
address = {In \textit{Afroasiatica} \textit{Italiana}, 171-188. Napoli},
author = {Blažek, Václav},
publisher = {Instituto Universitario Orientale},
sortname = {Blazek, Vaclav},
title = {Cushitic Lexicostatistics: {{T}}he second attempt},
year = {1997}
}


Blažek, Václav. ms [2000]. Fragment of a comparative and etymological dictionary of

@book{Blažek2003,
address = {In \textit{Das} \textit{alte} \textit{Ägypten} \textit{und} \textit{sei
author = {Blažek, Václav},
publisher = {Österreichisches Literaturforum},
sortname = {Blazek, Vaclav},
title = {Beja kinship and social terminology},
year = {2003a}
```

*??*

}

@book{Blažek2003,
address = {A fragment of a comparative-etymological dictionary of Beja. In \text
author = {Blažek, Václav},
publisher = {Russian State University for Humanities},
sortname = {Blazek, Vaclav},
title = {Fauna in Beja lexicon},
year = {2003b}
}

@article{Blažek2006,
author = {Blažek, Václav},
journal = {(Fragment of a comparative and etymological dictionary of Beja). \tex
pages = {365--407},
sortname = {Blazek, Vaclav},
title = {Natural phenomena, time and geographical terminology in Beja lexicon},
volume = {2},
year = {2006}
}

@book{Blažek2007,
address = {In \textit{Omotic} \textit{and} \textit{Cushitic} \textit{language} \
author = {Blažek, Václav},
publisher = {Rüdiger Köppe},
sortname = {Blazek, Vaclav},
title = {Beja historical phonology: {{C}}onsonantism},
year = {2007a}
}

@article{Blažek2007,
author = {Blažek, Václav},
journal = {(Fragment of a comparative and etymological dictionary of Beja)'. \te
pages = {383--428},

532

```
sortname = {Blazek, Vaclav},
title = {Natural phenomena, time and geographical terminology in Beja lexicon},
volume = {3},
year = {2007b}
}


@book{Browne2003,
address = {Champaign (ILL)},
author = {Browne, Gerald M.},
publisher = {Stipes},
title = {\textit{Textus} \textit{Blemmycus.} \textit{Aetatis} \textit{Christianae}
year = {2003}
}


@book{Cohen1984,
address = {Leuven, Paris},
author = {Cohen, David},
publisher = {Peeters},
title = {\textit{{La}} \textit{phrase} \textit{nominale} \textit{et} \textit{l'évo
year = {1984}
}


@book{Cohen1988,
address = {In \textit{Les} \textit{Langues} \textit{dans} \textit{le} \textit{Mond
author = {Cohen, David},
publisher = {Editions du CNRS},
title = {Bédja},
year = {1988}
}


@book{El-Sayed2011,
address = {\textit{Afrikanischstämmiger} \textit{Lehnwortschatz} \textit{im} \text
author = {El-Sayed, Rafed},
publisher = {Orientalia Lovanensia Analecta (OLA) 211},
```

*??*

year = {2011}
}


@book{HamidAhmed2005,
address = {Paris, Louvain},
author = {Hamid Ahmed, Mohamed-Tahir},
publisher = {Peeters},
title = {\textit{"Paroles} \textit{d'hommes} \textit{nonorables".} \textit{Essai
year = {2005a}
}


@book{HamidAhmed2005,
address = {In \textit{Land,} \textit{ethnicity} \textit{and} \textit{political}
author = {Hamid Ahmed, Mohamed-Tahir},
publisher = {Africa World Press/The Red Sea Press},
title = {Ethics and oral poetry in Beja society ({Sudan})},
year = {2005b}
}


@book{Hudson1976,
address = {In \textit{The} \textit{non-Semitic} \textit{languages} \textit{of} \
author = {Hudson, Richard A.},
publisher = {African Studies Center, Southern Illinois University},
title = {Beja},
year = {1976}
}


@book{Morin1995,
address = {Paris, Louvain},
author = {Morin, Didier},
publisher = {Peeters},
title = {\textit{"Des} \textit{paroles} \textit{douces} \textit{comme} \textit{l
year = {1995}
}

```
@misc{Onour2015,
author = {Onour, 'Abdallah},
note = {University of Khartoum.},
title = {Asbaab al-'ummiyya 'and al-Beǧa (Diraasaat ħaala, wilaayat Kasala, maħall
year = {2015}
}


@book{Reinisch1893,
address = {Wien},
author = {Reinisch, Leo},
note = {Akademie der Wissenschaften in Wien L, Band CXXVIII.},
publisher = {Sitzungs berichte der Kais},
title = {\textit{Die} \textit{Beḏauye-{Sprache}} \textit{in} \textit{Nordost-Afrika
year = {1893}
}


@book{Reinisch1895,
address = {Wien},
author = {Reinisch, Leo},
publisher = {Alfred Hölder},
title = {\textit{{Wörterbuch}} \textit{der} \textit{Beḏawiye-{Sprache}}},
year = {1895}
}


@book{Rilly2014,
address = {In \textit{The} \textit{Fourth} \textit{Cataract} \textit{and} \textit{
author = {Rilly, Claude},
publisher = {Peeters},
title = {Language and ethnicity in ancient {Sudan}},
year = {2014}
}
```

Rilly, Claude. forth. [2018]. Languages of Ancient Nubia. In \textit{Handbook} \

@book{Roper1928,
address = {Hertford},
author = {Roper, E. M.},
publisher = {Stephen Austin},
title = {\textit{Tu} \textit{Beɖawiɛ.} \textit{An} \textit{elementary} \textit{h
year = {1928}
}


@book{Ullendorf1955,
address = {\textit{The} \textit{Semitic} \textit{languages} \textit{of} \textit{
author = {Ullendorf, Edward},
note = {Taylor's.},
publisher = {London},
year = {1955}
}


@book{vanCoetsem1988,
address = {Dordrecht},
author = {van Coetsem, Frans},
publisher = {Foris},
title = {\textit{Loan} \textit{phonology} \textit{and} \textit{the} \textit{two}
year = {1988}
}


@book{Vanhove2003,
address = {In \textit{AIDA} \textit{5th} \textit{Conference} \textit{Proceedings
author = {Vanhove, Martine},
publisher = {Servisio de Publicaciones Universidad de Cádiz},
title = {Bilinguisme et alternance bedja-arabe au Soudan},
year = {2003}
}

@book{VanhoveHamid Ahmed2011,
address = {In \textit{Dictionnaire} \textit{des} \textit{langues}, eds. Emilio Bon
author = {Vanhove, Martine, and Hamid Ahmed, Mohamed-Tahir},
publisher = {PUF},
title = {{Le} bedja},
year = {2011}
}


@book{Vanhove2012,
address = {In \textit{Morphologies} \textit{in} \textit{contact}, eds. Martine Van
author = {Vanhove, Martine},
publisher = {Akademie Verlag},
title = {Roots and patterns in Beja (Cushitic): {{T}}he issue of language contact
year = {2012}
}


@book{Vanhove2016,
address = {In \textit{Finiteness} \textit{and} \textit{Nominalization}, eds. Claud
author = {Vanhove, Martine},
publisher = {Benjamins},
title = {The Manner converb in Beja (Cushitic) and its refinitization},
year = {2016}
}


@book{Vanhove2017,
address = {Leuven, Paris},
author = {Vanhove, Martine},
publisher = {Peeters},
title = {\textit{{Le}} \textit{bedja}},
year = {2017}
}


@book{Wedekind2012,
address = {In \textit{Proceedings} \textit{of} \textit{the} \textit{6th} \textit{W

*??*

author = {Wedekind, Klaus},
publisher = {Rüdiger Köppe},
title = {Sociolinguistics developments affecting Beja dialects},
year = {2012}
}


@book{Wohlgemuth2009,
address = {Berlin, New York},
author = {Wohlgemuth, Jan},
publisher = {Mouton de Gruyter},
title = {\textit{A} \textit{typology} \textit{of} \textit{verbal} \textit{borrow
year = {2009}
}


@book{Záhořík2007,
address = {In \textit{Beiträge} \textit{zur} \textit{1.} \textit{Kölner} \textit
author = {Záhořík, Jan},
publisher = {Accessed 28 \citealt{February2018}},
sortname = {Zahorik, Jan},
title = {The Islamization of the Beja until the 19th century},
year = {2007}
}


@book{Zibelius-Chen2014,
address = {\textit{Lingua} \textit{Aegyptia.} \textit{Journal} \textit{of} \text
author = {Zibelius-Chen, Karola},
publisher = {267-309},
title = {{Sprachen} Nubiens in pharaonischer {Zeit}},
year = {2014}
}

# Did you like this book?

This book was brought to you for free

# Arabic and contact-induced language change

change blurb in localmetadata.tex