

# LANGUAGE INTERRUPTED

Signs of Non-Native  
Acquisition in  
Standard Language  
Grammars

John McWhorter

# Language Interrupted

*This page intentionally left blank*

# Language Interrupted

---

*Signs of Non-Native Acquisition  
in Standard Language Grammars*

JOHN McWHORTER

**OXFORD**  
UNIVERSITY PRESS

2007

# OXFORD

UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further  
Oxford University's objective of excellence  
in research, scholarship, and education.

Oxford New York  
Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in  
Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2007 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.  
198 Madison Avenue, New York, New York 10016

[www.oup.com](http://www.oup.com)

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, recording, or otherwise,  
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

McWhorter, John H.

Language interrupted: signs of non-native acquisition in standard language grammars /  
John McWhorter.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-19-530980-5

1. Typology (Linguistics). 2. Linguistic change. 3. Languages in contact. I. Title.

P204.M34 2007

415—dc22 2006051521

9 8 7 6 5 4 3 2 1

Printed in the United States of America  
on acid-free paper

# Acknowledgments

I could not have written this book without the assistance of various linguists who were kind enough to apply their expertise to providing me with either crucial data or informed counsel to assist me in pronouncing upon language families with which I had never before had occasion to become familiar.

For the English chapter, sincere thanks to Andrew Garrett for extensive comments on an early draft of the article that preceded it, to Osten Dahl for reining me in on some of my preliminary claims, and otherwise to Irmengard Rauch, Anthony Grant, Gary Holland, James Matisoff, Peter Tiersma, Jarich Hoekstra, Martin Haspelmath, Werner Abraham, Sally Thomason, Elly van Gelderen, and especially my once and future mentor Elizabeth Traugott.

For the Chinese chapter, Hilary Chappell and Stephen Matthews went far beyond the call of duty in helping me penetrate the Sinitic literature beyond the easily available sources on Mandarin. Professor Chappell was especially invaluable in providing me with data translated from the Chinese as well as work in progress that helped me get a grasp on comparative perspectives unavailable elsewhere.

Gernot Windfuhr and Prods Skjaervø kindly assisted me in filling in some gaps for the Persian chapter, and Jay Jasanoff gave me an invaluable crash course on comparative Indo-European diachronic morphology, especially useful for assessing Persian in comparison to its sisters.

Without Alan Kaye and his innately comparative perspective that jibes with my own, I would be unable to feel confident enough to present the Arabic chapter for public assessment. His generosity in providing endless offprints and free refereeing is hereby acknowledged in awe and gratitude.

For the Malay chapter, I owe endless thanks to Austronesianists Robert Blust, Mark Donohue, and John Woolf for references and insights that helped in constructing the argument. I am also deeply grateful to David Gil for sharing his work on Riau Indonesian with me, which helped to stimulate my interest in the topic of this monograph and also served as my introduction to Indonesian languages. Gil also kindly pointed me to Lea Brown's dissertation on Nias. A special acknowledgment is also due to

Geoffrey Hull, who provided me with invaluable discussion of the Timor situation and steered me to sources that were unlikely to come to my attention otherwise.

Presentations of various stages of this work to audiences at the University of California, Berkeley, the University of Helsinki, the University of Pennsylvania, Stanford University, and the Free University of Berlin also helped me to sharpen my argumentation.

Finally, I picked up the basic idea that language family members that retain most of their ancestors' complexities are "normal" from work by, and constant exchanges with, Stephane Goyette. If I had never encountered his argument that Romance is not a "natural" outcome of Latin given the "conservative" nature of Slavic, Greek, and Celtic, I would likely never have written this book.

# Contents

1. Introduction	3
2. Defining Grammatical Complexity	21
3. Epistemological Caveats	51
4. English	59
5. Mandarin Chinese	104
6. Persian	138
7. Colloquial Arabic	165
8. Malay	197
9. A New Typology of Language Contact	252
Notes	277
References	283
Index	307



*This page intentionally left blank*

# Language Interrupted

*This page intentionally left blank*

# Introduction

## *Navajo as Default*

### 1. The Problem

The concept elicits little objection when observed in passing. Dalby (1998, 391) hardly considers it a venturesome statement that “a lingua franca needs to be easy to grasp, and Malay has a more approachable structure than its relatives.” When Trudgill (1996, 8–9) proposes that Alpine Romansch and Faroese have accreted unusual and complex sounds because as isolated languages, they have rarely been subjected to adults’ diminished capacity for imitating unfamiliar phonetic segments, the argument is not received as controversial.

That is, that heavy second-language acquisition decreases structural complexity is thoroughly intuitive to most linguists on a certain level. Yet in response to a specific argument that non-native acquisition has rendered a given language less complex overall, linguists tend toward a visceral skepticism. This skepticism is founded upon an assumption considered a hallmark of informed linguistic thought, that human grammars are all equally complex. Some examples of this commonplace assumption among linguists include Edwards (1994, 90), Bickerton (1995, 67), O’Grady et al. (1997, 6), and Crystal (1987, 6–7). We are to keep in mind, for example, that English may appear a “simplified” Germanic language in lacking Icelandic’s inflectional paradigms, but has *do*-support, and a mixed phonology with different rules for Germanic words than Latinate ones.

Responses to complexity arguments are especially vehement when it is proposed that a creole language is less complex than older languages. Reactions tend to fall upon a cline ranging from sharp objections suspecting sinister sociopolitical motives (various writings of Michel DeGraff) to studious disinterest (e.g., Roberge 2003, 343: “I will not be losing any sleep over the coherence of the field if creoles are not a linguistically definable class”).

But even proposals that non-native acquisition has simplified older, non-creole languages tend to meet a resistance traceable to the guiding sense that there is no such thing as grammarwide simplification to any significant degree. Many writers

point to the fact that inflectional losses that may appear to suggest interrupted transmission were already underway before the language entered the contact situation in question (e.g., Thomason and Kaufman 1988, 263–342 on English; or Corriente 1971; Diem 1973; and Zwettler 1978 on Arabic). The implication, then, is that the simplification was merely that which is ordinary as grammars change internally.

Then also, language contact work tends strongly to focus on mixture rather than reduction. This is based on a tacit assumption that because no language is ever less complex than another, when Language A is altered in contact with Language B, then Language A undergoes structural replacements and alterations based on Language B, but not abbreviations, unknown in either language, that are due to speakers of Language B imperfectly mastering Language A.

If a difference in overall complexity between an older language and its relatives remains stark enough to tempt an explanation based on non-native acquisition, then the paradigm tends to force the analyst into appealing to a type of language all agree is less complex than real ones: pidgins. Therefore many accounts propose that a suspiciously simplified older language was dissolved to the extent of pidginization at some point. Examples include Bailey and Maroldt's (1977) argument that English was creolized in contact with Norman French, and Versteegh's (1984) argument that the regional varieties of Arabic began as a pidgin Arabic. But the absence of useful documentation of the pidgin in question regularly bars these accounts from prevailing. Other specialists deign it more graceful to return to arguments referring to mixture or internally driven inflectional loss (e.g., Diem [1978] preferring to seek signs of mixture in Arabic; Thomason and Kaufman [1988, 263–342] on inflectional collapse in English before the Norman takeover).

## 2. The Response

This monograph is an argument that this neglect of the role of reduction in language contact beyond pidginization constitutes an empirical gap, neglecting a commonly encountered qualitative difference between related languages. Our models of language contact are weakened to the extent that they squarely address the effects of adult acquisition only in pidgins, but otherwise acknowledge them only in vague statements about a given language being streamlined because of its *lingua franca* status, or in treatments of fragments of a language's structure such as a few marked sounds.

In attempting to fill in the empirical gap in question, I have elsewhere argued extensively that creoles, despite being full languages, have less complex grammars than older ones', such that creoles represent an extreme along a cline of structural reduction that a language may undergo in a contact situation (cf. McWhorter 2001, 2004). In this book, my interest is in filling in the middle range of the gap. Here fall a number of languages, many standardized and vehicles of literatures, whose timelines include periods during which I will argue that adult acquisition abbreviated the complexity unnecessary to communication that accretes in all grammars when they are passed down generations in complete form.

This book is constructed to advance the following argument: in the uninterrupted transmission of a human language, radical loss of complexity throughout the gram-

mar is neither normal, occasional, nor rare, but *impossible*. The natural state of human language is one saddled with accreted complexity unnecessary to communication. Wherever this complexity is radically abbreviated overall rather than in scattered, local fashion, this is not just sometimes, but *always* caused by a sociohistorical situation in which non-native acquisition of the language was widespread enough that grammar was transmitted to new generations in a significantly simplified form. This is true not only in the extreme case of pidgins and creoles, but also to a lesser but robust extent in many languages of the world.

My argument will be founded on three basic assumptions.

### 3. Assumption One: In Ordinary Language Change, Grammars Never Become Radically Less Complex Overall

#### 3.1. *Simplification as a Major Player*

In much academic discussion of how grammars change, there is an implication that it is ordinary for a grammar, or a significant portion of it, to radically simplify in comparison to its close relatives, with the simplification not conditioning new complexifications in its wake. Thus while it has long been known that grammars both simplify and complexify (e.g., Paul 1880), it is assumed that even under ordinary conditions of transmission, simplification can proceed to a striking degree.

Certainly a great deal of discussion depicts replacement rather than simplification. For example, grammaticalization, in which lexical morphemes become grammatical or pragmatic ones, entails reshuffling that neither simplifies nor complexifies. The evolution in Ewe of nominal *megbé* “back (of a person)” into an adposition meaning “behind” or “after” (Heine, Claudi, and Hünnemeyer 1991) is simply a semantic transformation: there are no grounds for treating adpositions as more complex than nouns.

All analysts are also aware that grammars complexify, such as the development in English of *do*-support.

But simplification is considered an equally likely pathway, and not just in terms of, for instance, local sound mergers and their moderate effects upon morphology, but in terms of entire structural modules. It is assumed that it is normal that a grammar might, for example, shed vast amounts of its morphology without replacing it with new morphemes, such as English in contrast to Icelandic. Persian’s loss of ergativity is treated as a mere matter of chance, despite such a development being almost unknown in other Indo-Iranian languages beyond a few of Persian’s closest relatives in long-term intimate contact with it. Middle Chinese had four tones in two register (phonation) distinctions. In Mandarin this devolved into four tones with no register distinctions, mediated by a small number of sandhi rules. In all the other modern Chinese dialects, however, there are six to nine lexical tones, often with sandhi rules more complex than Mandarin’s, even extending to lexical specification as in Min (Taiwanese and Fujianese) and Wu (Shanghainese). Yet the Mandarin development is treated as a mere roll of the dice.

That is, it is supposed that just as a grammar might gradually *replace* focus-marking machinery with conjugational paradigms and/or voice-marking machinery

as in Austronesian, or might *develop* a new subjunctive paradigm as Modern Aramaic did, it is an equally likely possibility that it *sheds* all marking of grammatical gender as English did (uniquely, it must be noted, among all Indo-European languages of Europe).

### 3.2. *The European Problem: The “Drift” to Analyticity*

Indeed, the sense that grammars naturally “take it all off” in entire modules of grammar is encouraged by cases of the English type. The emergence and societal entrenchment of printing in Europe and its spread to the Middle East meant that the changes from Proto-Germanic to English, Latin to Romance, Old Persian to Modern, and so forth were documented in writing. Then the academic culture that printing and literacy allowed meant that centuries later, most academic linguists would be native speakers of these languages. Enter, then, a tradition in which central points of reference are grammars that abbreviated or eliminated distinctions of grammatical gender, definiteness, voice and other features; replaced many affixal constructions with analytic ones; and erased vast degrees of suppletivity and irregularity. Resolutely synthetic exceptions like Icelandic are treated as “conservative” outliers.

This is commonly described as an Indo-European family “drift” to analyticity, presumably a thoroughly ordinary aspect of how grammars might change. To the extent that it is considered anomalous, this is in its occurrence across a broad area rather than in its development within an individual grammar. Hence an assumption that radical simplification in, for example, the vast realm of inflectional morphology is one path a grammar might take.

To be sure, analyticity is not simpler than boundedness in the necessary sense. A bound morpheme (e.g., past marker *-ta* in Japanese *ki-ta* “wore”) is not “more complex” than a free one of the same meaning and function (e.g., *dây* in Thai, as in *dây pay* “went”). But the analyticity that came to prevail in some Indo-European families did, as often as not, simplify rather than replace. First, semantic distinctions were often eliminated entirely rather than replaced; that is, Romance languages do not replace Latin’s neuter gender marking with a free morpheme. Second, when affixal distinctions are indeed retained via free morphemes, such as Romance’s indication of the genitive with an adposition such as French *de* rather than a case ending, this minimizes the complications occasioned by allomorphy (e.g., Latin genitive marking varied across declension classes: *vi-ae* “of the road,” *popul-i* “of the people,” *lēg-īs* “of the law”) and morphophonemics due to the exigencies of long-term adjacency of segments (nominative *honōs*, but genitive *honōr-is*). Thus because inflectional loss typically entails simplification, to assume that the analytic “drift” in Indo-European was a normal internal development is indeed to suppose that massive simplification across the morphological module of grammars is unremarkable.

The Indo-European case has even led to a more general assumption: that analytic “drift” is a cross-linguistic tendency, and that in general, a tendency toward simplification is inherent to grammars worldwide. Languages elsewhere where no such “drift” is evident then appear exceptional. Franz Boas (1938, 160) indicated the mindset in question:

Many primitive languages are complex. Minute differences in point of view are given expression by means of grammatical forms; and the grammatical categories of Latin, and still more so those of modern English, seem crude when compared to the complexity of psychological or logical forms which primitive languages recognize, but which in our speech are disregarded. On the whole, the development of language seems to be such, that the nicer distinctions are eliminated, and that it begins with complex and ends with simpler forms, although it must be acknowledged that opposite tendencies are not by any means absent.

Note “not by any means absent” —implying that such “opposite tendencies” are mere static, amid a general “drift” toward simplification.

But whence the prominence Boas attributes to simplification, if the “primitive” languages he refers to do not exhibit it? Despite his pioneering and invaluable commitment to dispelling conceptions of European languages as more “advanced” than indigenous ones, Boas’s assumption was based on the immediacy of the documentation of Indo-European languages, in which inflectional loss is so often evident.

### 3.3. *Around the World with a Typologist*

This is thoroughly understandable, but that documentation covers at most the past three millennia. We might ask: if human language has existed for at most 150,000 years and according to the most conservative estimates a few tens of thousands, then exactly why would this presumably worldwide “drift” to simplicity have only kicked in over the past two millennia or so?

In this light, across the globe Boas’s “primitive” languages vastly outnumber the two hundred-odd languages out of six thousand that have a substantial written tradition. And a bird’s-eye view—or if I may, “typologist’s-eye view”—of the world’s languages renders questionable an assumption that abbreviation plays a dominant role in how human language grammars develop.

We might literally imagine the world’s languages as they cover the globe. The indigenous languages of North America are highly inflected, with inflections marking categories beyond “Standard Average European” ones such as obviativeness and the inverse, and with boundedness often so central to the grammars as to occasion polysyntheticity and noun incorporation. The indigenous languages of South America are similarly elaborated grammatically, commonly having paradigms of evidential markers and noun classifiers. Flying due southwest to Australia, we encounter hundreds more languages bristling with inflections, these often marking split ergativity, and beyond this, noun classes, distinctions of number and inclusivity within plural pronouns, and so on.

Importantly, there are no indigenous languages of the Americas or Australia that have taken the supposedly natural route of “drifting” into complete or even near-complete analyticity, or of shedding so many semantic and syntactic distinctions that they would suggest prior pidginization or creolization. Proto-Algonquian, for example, is reconstructed as a highly inflected language giving no indication under any analysis as “simplified,” and its modern descendants all submit to the same description.



No Algonquian language contrasts with the reconstructed proto-language as English does with Proto-Germanic.

Then we fly on to Asia. Its Austronesian languages almost all have either elaborate focus marking like the languages of the Philippines or conjugational paradigms reminiscent of Romance. Languages of the Southeast Asian Sprachbund (e.g., Sino-Tibetan, Mon-Khmer, Tai-Kadai, Miao-Yao) tend to have several lexical tones, vast proliferations of vowels, and/or batteries of numeral classifiers. The Semitic languages have their famous nonconcatenative morphology with semantics and syntax encoded via affixal and vocalic alternations amid trilateral consonantal roots, the systems riddled with irregularities. Then in Africa, most Niger-Congo languages have multiple noun classes; within this family, West Atlantic has awesomely complex consonant mutations, and Kwa and close relatives have complex tone systems. Nilo-Saharan languages offer assorted complexities of tone and inflection. No Khoi-San language has mysteriously eliminated the family's click sounds and the notoriously vast phonemic inventories these tend to occasion, and many are well inflected.

And then we return to Europe. The consonant gradations alone in Uralic languages like Finnish disqualify them immediately as having "drifted" toward any kind of simplicity. And even in Indo-European, Romance and Germanic are, after all, but two subfamilies out of nine. Equally representative are Baltic, Slavic, Greek, and Celtic, which, if they were the only surviving branches of the family, would have occasioned no observation such as Boas's.

What this flyover of the world shows is that in general, languages tend very strongly to exhibit high degrees of elaboration in at least some portion of their structures. If there is not heavy inflection, there are likely many tones and the sandhi rules they occasion. Here is found ergativity and switch-reference markers, there is found consonant mutations and noun classes, and so on.

There is always something—and there is no evidence that grammars regularly shed these complexities for some reason as a matter of course. Rather, we see grammars holding on to such things for vast periods of time. Dahl (2004, 269–74), for example, points out that the irregular past-marking ablaut patterns in Germanic strong verbs have lasted for millennia in all of the languages of the family but Afrikaans since Proto-Germanic, with languages even innovating new ones as some older ones fall away. He in the same vein notes that Semitic nonconcatenative morphology—alive and well in all of the family's members—traces back five thousand years at the very least according to the concrete documentation of Akkadian, and via reconstruction, likely traces to the birth of Proto-Afro-Asiatic *ten* thousand years ago. In human language grammars, complexity is ordinary, robust, and tenacious.

### 3.4. *Simplification Nibbles at the Margins*

Certainly simplification occurs in various parts of a grammar. For example, local collapses in inflectional paradigms are commonplace in language change. But our world tour suggests that it is hardly the usual case that this kind of change snowballs, conditioning a "drift" or "conspiracy" that leaves a language virtually analytic.

For example, inflectional erosion is hardly unknown in Australia, but it occurs within languages that remain highly synthetic. In the Mayali dialect chain, Kunwinjku

preserves four-form gender agreement, while Kune has extended the masculine agreement marker to all nouns. But the nouns themselves remain distinguished by class (Evans, Brown, and Corbett 2002, 117):

TABLE 1.1. Noun Classes and Agreement Markers in Mayali Dialects (Kw. = Kunwinjku; Kn. = Kune)

<i>good boy</i>	<i>good old.woman</i>	<i>good food</i>	<i>good rock</i>
Kw. na-rangem na-mak	ngal-kohbanj ngal-mak	man-me man-mak	kun-wardde kun-mak
Kn. na-rangem <b>na</b> -mak	ngal-kohbanj <b>na</b> -mak	man-me <b>na</b> -mak	kun-wardde <b>na</b> -mak

Crowley (2000) notes that the southern Vanuatu Oceanic language Ura, despite having never been a lingua franca, is somewhat simpler inflectionally than its relative Sye, in terms of degree of allomorphy and depth of morphophonemics. Differences in degree of this kind are inevitable—but Ura remains very much an inflected language.

Moreover, quite often simplification occasions new complexity in its wake. This is well known in the operations of analogy, such as in Old English, in which a morphophonemic rule in strong verbs that rendered [s] as [r] before plural past markers and the past participial ending unraveled to various degrees in verbs (such as before the past participle with *wesan* “to be” and before both the participle and the past plural ending with *lesan* “to read”), leaving an irregular situation (Hock 1991, 259):

TABLE 1.2. Variable Dissolution of [s] > [r]  
Morphophonemic Rule in Old English

infinitive	plural past	past participle	
<i>cēosan</i>	<i>curon</i>	(ge)coren	choose
<i>wesan</i>	<i>wæron</i>	(ge)wesen	be
<i>lesan</i>	<i>læson</i>	(ge)lesen	read

But such complementary processes also occur beyond the realm of local analogies. Hyman (Forthcoming) argues that the agglutinative structure of most Bantu languages did not result from the coalescence of the isolating structure of Niger-Congo relatives such as the Kwa group, but rather that the agglutinative typology was the original situation. In languages like Twi and Yoruba, the breakdown of the agglutination decreased complexity, as the noun class markers were fossilized or disappeared, and the morphophonemic processes caused by boundedness disappeared. But then tone was left with a greater contrastive role on the resultant monosyllables, developing attendant processes of sandhi. Mon-Khmer languages have often lost tone and/or register distinctions, but then these leave behind some of the vastest inventories of vowels in the world (often as many as three dozen and more).

Then there are cases in which new complexity is unconnected to simplification. English’s *do*-support was unconnected to its loss of grammatical gender. The linking in Slavic of animacy to case, in which animates mark the accusative with the genitive

(*Ivan ubyl brat-a* “Ivan killed the brother”) was unrelated to, for example, the family’s peculiar elimination of passive and subjunctive morphology. Romance languages lost most or all of Latin’s nominal case marking, neuter gender, and inflectional paradigms for passive and middle voice, but meanwhile innovated definite and indefinite articles, an animacy distinction in pronouns, deictic object pronouns, a conditional paradigm, and larger phonemic inventories encompassing nasals (French, Portuguese), high central vowels (Romanian), or front rounded ones (e.g., French, some Rhaeto-Romance varieties).

Thus the general tendency is that grammars maintain a high level of complexity. Simplification occurs, but is counterbalanced by new emergences. There are a vast number of semantic distinctions that a grammar may wend into marking; a wide variety of sounds that a grammar may develop; and a wide array of syntactic configurations that a grammar may take on. It is inherent to language change to nose into a healthy subset of the potential total of these distinctions available to human language. This process is inexorable, and when passed on uninterruptedly across generations, grammars do not mysteriously detour from it. Simplification nibbles at the margins.

Here, then, is where scholars founder when they dissociate curiously extensive simplification in a language’s history from a contact situation by referring to scattered inflectional elisions at an earlier period. Examples include Thomason and Kaufman (1988, 263–342) on English; Naro and Scherre (2000) on Popular Brazilian Portuguese; Kloecke (1950) and Bosman (1962) on Afrikaans; Chaudenson (1992) on Mauritian Creole French; and Corriente (1971), Diem (1973), and Zwettler (1978) on Arabic. It would be strange not to find such elisions at any stage of a grammar’s history: even the most conservative Indo-European languages collapsed or eliminated various inflectional distinctions in the proto-language. The issue is actualization: that is, just why simplification (including that beyond the inflectional, an issue usually marginalized in works such as those listed above) was so especially rampant after the contact period.

Where a grammar is markedly simplified in many areas compared to its close relatives this is a signal that something irregular has occurred. That is, my assumption in this book is that in terms of the quizzical contrasts between English and the other Germanic languages, or Mandarin and its sister Chinese languages, where there’s smoke, there’s fire.

### 3.5. *Darwin Cannot Help Us Here*

In this, the frequent analogy of language change with biological evolution breaks down. Animals have become more complex over time, with one-celled organisms evolving into mammals such as ourselves. But as Gould (1996) argues, increasing complexity is but an option in natural selection. Countless creatures have remained primordially simple for billions of years, such that, properly speaking, bacteria in their vast number and awesome eternity over the history of the planet are the most successful organisms on earth. Gould (149–51) analogizes the optional emergence of complexity to a drunkard staggering between the wall of a bar and a gutter where he risks falling into a stupor. Which direction he will stagger in at each step is unpredictable, and yet he will always eventually wind up passed out in the gutter — because

if by chance he does stumble into it, he will fall and no longer be able to continue. Thus organisms “stumble” into complexity rather than being driven toward it. Some will inevitably end up passed out in the gutter of complexity, but just as many will bounce against the wall and keep staggering on in glorious simplicity.

Despite controversy as to Gould’s claim that there is no directionality in evolution (cf. Morris 2003), it is indisputable that a great many creatures remain quite simple indefinitely. However, in this book my claim is that this is impossible of grammars. Once human language arose, all grammars eventually achieved a high degree of complexity. None randomly remained unadorned with anything but that which is necessary to communication: there are linguistic analogues neither to bacteria nor the wall that the drunkard may lurch against. Then, short of an intervention of some kind, all grammars remained highly complex *in perpetuo*. No degree of ordinary simplification returns grammars to a “bacterial” state. To the extent that grammatical simplification in a language would appear to have reverted a primate into a reptile, so to speak, this suggests that there has been an interruption in the regular transformation of the grammar.

### 3.6. *Esoteric versus Exoteric Languages*

Thus this argument proceeds upon Thurston’s (1987) distinction between “esoteric” and “exoteric” grammars. In esoteric grammars, because they are rarely acquired by non-native speakers, sound change is free to create allomorphy and suppletivity, materials are recruited to mark ever more fine semantic distinctions, and that which is optionally marked is especially likely to become regularized. Exoteric grammars are those that have been subjected to extensive non-native acquisition. Adults, with their ossified capacity for acquiring languages, tend to substitute sounds harder to acquire, simplify morphological opacities, and eliminate semantic distinctions and syntactic constructions tangential to conveying urgent aspects of meaning or that are difficult to glean for the non-native. The implication of Thurston’s argument is that “esoteric” languages are a natural state.

Perkins (1992) makes a related argument that, in languages used by small groups, shared immediate context encourages the overt marking of fine distinctions readily apparent to all members that would be less apparent to outsiders. He shows that the presence and degree of features such as person affixes, dual marking, inclusive/exclusive pronominal distinctions, tense marking, gender marking, and the degree to which demonstratives and adverbial deictics carve up semantic space are predictable according to what Thurston would term the esoteric/exoteric distinction. Thus !Kung, Andamanese, Yupik, and Ojibwa were among the highest-ranking languages on this scale.<sup>1</sup> Such languages retain these distinctions over time despite the ravages of sound change. Heath (1997) associates this tendency to a stable sociolinguistic environment, as opposed to one in which transmission is disrupted by non-natives.

### 3.7. *What Is a “Normal” Language?*

Thus my assumption is that the “ordinary” language is one like the Nakh-Daghestanian language Chechen:

- (1) Maliikina Ahwmad ‘a gina, cunna tuoxa ‘a toexna,  
 Malika.DAT Ahmed CONN see.WP 3S.DAT hit.INF CONN hit.CVant  
 cynga shiena ‘a tuoxiitira.  
 3S.LOC 3S.RFL.DAT CONN hit.CAUS.WP

‘Malika saw Ahmed, she hit him, and then he hit her.’ (Good 2003a)

(CONN = connective, WP = witnessed past, Cvant = anterior-marking converb,  
 RFL = reflexive, CAUS = causative)

Here, the third *she* is marked as reflexive, as a grammaticalized strategy for long-distance anaphoric marking. This is a sign of significant bleaching and reanalysis of the core meaning of reflexivity. The final *hit* is marked as causative to indicate that its oblique argument is coreferential with the subject of the preceding clause. There are no causative semantics in the latter case; obviously its use as a switch-reference marker indicates drift beyond any recoverable connection with its core meaning. But then, because causatives occasion that their object be marked with the locative case, the third-person object pronoun in the third clause is locative marked. Meanwhile, Chechen has dative-marked subjects, and in clause chains, the dative marking is carried over onto the subjects of the following clauses, such that the third *she* is marked both as reflexive and as dative. The reflexive pronoun is a suppletive form. Chechen has grammaticalized evidential marking, as we see with the witnessed past marking. Verb stems vary for present, past, and the infinitive through forms of ablaut. Such things are evidence of a grammar’s having existed for a very long time used mainly by native speakers.

Creole languages, on the other hand, are born in contexts in which non-native acquisition is a norm. This produces renditions of natural language much less complex than Chechen. In Saramaccan, the Chechen sentence is:

- (2) Malíka bi sí Améd, a fómẽẽ, hẽ Malíka a fómẽẽ.  
 Malika PAST see Ahmed, 3S hit.3SOBJ then Malika 3S hit.3SOBJ

‘Malika saw Ahmed, she hit him, and then he hit her.’ (Good 2001)

Saramaccan is in no sense devoid of its own complexities. It distinguishes case in the third-person singular pronoun (and also the second-person plural one). In the form *fómẽẽ* the oblique third-person singular pronominal has cliticized, altered morphophonemically (but only from *ẽ* to *ẽẽ*), and is preceded by a morphophonemically conditioned [m] (which, however, only occurs with this verb and *njã* ‘to eat’). Not indicated is that tone sandhi spreads high tone to the right but breaks at left edge maximal projections, such as between verb and object, and thus [*\*sí Áméd*]. However, in terms of complexity, a difference of degree between Chechen and Saramaccan is clear. It is a symptom of the fact that Saramaccan, like all creoles, was born amid rapid non-native acquisition, in its case that of English and Portuguese by West Africans. This book will explore cases in between extremes like Chechen and Saramaccan.

My conception of language change entails, then, that languages like Chechen are the normal state of a grammar. But the attractions of the linguistic relativism frame-

work, founded upon the Sapir-Whorf hypothesis, may condition a sense that we are faced instead with a purely synchronic cultural difference, unconnected with any ontological distinction between the normal and the derived, or the primary and the secondary.

Along these lines, one would suppose that in small societies, the intimacy between members sparks the semantic overspecifications of a language like Chechen or Kunwinjku, and that the lesser degree of this overspecification in a lingua franca like English is due to a lesser degree of such intimacy. But this perspective might entail no implication that the English type of grammar arose from the abbreviation of one more like Chechen's. One might even reverse the developmental relationship I propose and assume that the English type of grammar is the "default," with the Chechen type an elaboration upon this.

But there are three flaws with this perspective.

First, while it accounts for overspecification, it does not explain, in the strict sense, why smaller languages would tend to have larger and more marked phonemic inventories, or more irregularity and suppletivity. These are driven not by culture, but by the mechanical factor of long-term use of a conglomeration of sounds and grammatical constructions by native speakers.

Second, it leaves unexplained the many languages spoken by small groups and rarely learned by outsiders that are strikingly less accreted with the features Perkins treats than other languages: creoles. Languages like Saramaccan of the Surinamese rain forest or the Portuguese creoles of the Gulf of Guinea, spoken on tiny islands and largely unheard of by anyone beyond them, are acquired non-natively no more than obscure Caucasian or Amazonian languages. If the correlation between grammar and cultural type is a mere present-day matter, then these languages and many other creoles stand as glaring exceptions. Despite being today esoteric languages according to Thurston's taxonomy, creoles have relatively simple grammars because they were built up from pidgin-level competence just a few centuries ago. The deciding factor in grammatical complexity, then, is non-native acquisition, not present-day culture. The currently exoteric language, affected by non-native acquisition, will always be relatively simplified. But the currently esoteric language, despite being rarely learned non-natively, can be either simplified like Saramaccan or massively complex like Chechen, according to whether there was a period of extensive adult acquisition in its history.

Third, this perspective is a presentist one, neglecting the history of humankind. For most of human history, our species consisted of hundreds of thousands of small bands of hunter-gatherers. Somewhere during this period was when, by all accounts, the human language capacity emerged. Only with the development of agriculture did some groups begin expanding across vast territories, co-opting and exterminating original inhabitants and subsuming diverse peoples first into cities and later into cross-continental empires. Thus the post-Neolithic revolution exposed human language to new exigencies; namely, widespread non-native acquisition. This means that one type of grammar indeed came first and was the only kind for tens of thousands of years—and that was not the English type but the Chechen type.

Therefore, if we must view the difference between English and Kunwinjku as a cultural one, we must be aware that we underanalyze the cultural contrast if we view it only in the present tense, as if we were dealing with the difference between the Sardinian and the Serb. The cultural difference is due to a decisive *historical* transformation that humankind has undergone, in which the cultural context that Modern English arose in was a new development, while the one that Kunwinjku still exists in was original. English is the anomaly; Kunwinjku is bread and butter. Dahl's observation (2004, 295) is useful here:

The incidence of disruptive language contact and suboptimal transmission may well have increased historically. Certainly the average language of pre-agricultural humankind was low-contact, compared to the languages that most people speak today, or even in traditional agricultural societies. Languages spoken by small, nomadic groups in areas with a population density of about 1 person per square kilometre are unlikely to be exposed to large-scale suboptimal transmission.

### 3.8. *Changing the Lens*

This means that to propose that "indigenous languages tend to develop heavy degrees of elaboration" is an after-the-fact distortion, because heavy elaboration is inherent to human language. Rather, what merits remark is that "languages widely acquired non-natively are shorn of much of their natural elaboration."

When a ball falls off of a table, our natural tendency is to process this as the exceptional case, its staying in place having been "normal." But the realities of physics reveal this as an illusion. The laws of attraction entail the operations of gravity as the eternal condition, enforcing endless movement checked only by collision with another body. Only the intervention of the table holds the ball still, and only the floor keeps it from hurtling farther on. In language change, the development of needless complexity is a law like gravity; its abbreviation by non-native acquirers is an intervention like tables and floors.

Thus Bichlmeier and Marti (2003, 469) pose as a question why Slavic languages, despite sound change typically eroding local morphological distinctions in the grammars, did not lose their nominal morphology as some other Indo-European languages did. But, in fact, the question they might have asked is why the Romance languages *did* lose their nominal morphology to such a degree. That is, Bichlmeier and Marti's sense of what is ordinary is distorted. They are asking why the ball fell, when the real question is why the ball ever came to rest in the first place.

Given that proteins on cell membranes' surfaces play a central role in how cells recognize one another, one would naturally assume that phagocyte white blood cells would be stimulated to attack and eat a cell when the invader cell has a foreign protein on its membrane. But in fact, cellular biologists have discovered that often what stimulates a phagocyte to attack is that the invader cell *lacks* a protein that the phagocyte has. We ask "what makes the phagocyte eat a cell?" and suppose that it must be some foreign presence, when in fact it is an absence that sparks the behavior.

This polarity between presence and absence similarly determined that Polish is more elaborated than English, except in reverse fashion. Slavic languages are more elaborated than English not because of some peculiar *present* aspect of Slavic that left English unaffected, but because of something *present* in English that was absent in Polish—and Chechen and Kunwinjku. Polish has developed unimpeded; someone put their foot out and tripped English. The human grammar is a fecund weed, like grass. Languages like English, Persian, and Mandarin Chinese are mowed lawns, indicative of an interruption in natural proliferation.

On English, Sapir (1921, 170), in his exploration of the “drift” in English toward analyticity and eliminating fine distinctions like that between *here* and *hither*, gropingly intuited that the difference between English and relatives like German was due to an external factor. “Has the mechanical imposition of a flood of French and Latin loan-words, unrooted in our earlier tradition, so dulled our feeling for the possibilities of our native resources that we are allowing these to shrink by default?” he asked. I would specify that the dulled feeling was among Scandinavian Vikings learning English as a second language, but the spirit of Sapir’s surmise is congruent with the foundations of this book. For English to become the only Germanic language to entirely shed words like *hither*, *thither*, and *whither*, and so much else, cannot be treated as a fortuitous local peculiarity; it strongly suggests that some kind of external interference was at work.

#### 4. Assumption Two: Radical Simplification Can Be as Central in the Result of Language Contact as Mixture

##### 4.1. *Simplification as a Background Factor*

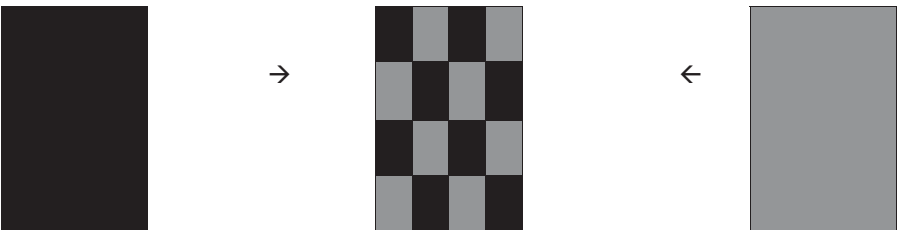
Work on language contact is generally based on a tacit assumption that when grammars come together, simplification is just one player among many, only dominating in the formation of pidgins and creoles. We list complexifications, such as the loss of word-final stops in the Nilo-Saharan language Baale because of contact with relatives that lack such stops, where, because the stops are retained before suffixes, the result is morphophonemic rules that did not exist before: *mèèlélé* “axe,” *mèèlélé-k-ká* “axes” (Dimmendaal 2001, 362). We list replacements, such as Ethio-Semitic languages’ converbial constructions modeled on Cushitic, replacing the finite coordinate and subordinate constructions (Thomason and Kaufman 1988, 135). Simplification, such as the loss of dual marking in Ethio-Semitic, is treated as just one more on the list. And often, the simplifications are treated as being modeled on the absence of features in the language encountered—such as the absence of dual marking in Cushitic—such that technically the simplification becomes a mere subtype of replacement.

The guiding schema behind this orientation is that when grammar A is affected by grammar B, the new version of grammar A will differ from the original one largely in having replaced some or many of its features with foreign ones. Here will be some complexification, there will be a bit of simplification, but the general sense appears to be that language contact is a kind of hybridization that leaves overall complexity intact:

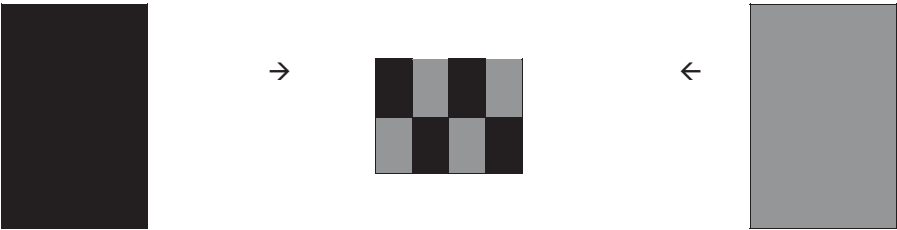




This would represent, for example, the effect of Slavic languages on Romanian. In cases where the two grammars make equal contributions, like *Media Lengua* with Spanish lexicon and Quechua phonology and morphosyntax, then the schema would be as follows:



Creolization, under an assumption that it results in languages less complex than older ones, involves languages mixing, but with all source language structures rendered in an abbreviated form (most explicitly argued by myself in various articles and by Keesing [1988], and implied by various articles on creoles drawing attention to the process, albeit without a commitment to an argument re overall complexity; cf. Siegel 2000a).



But none of these schemas capture a different kind of outcome of language contact. They miss that because language contact can entail degrees of incomplete acquisition lesser than pidginization, but significant nonetheless. This includes

simplification that is more pervasive than local wrinkles in the grammar, and is not modeled on features *present or absent* in Language B. These types of situations can be represented thus:



Thus when Scandinavians invaded England, the result was not a hybrid between English and Old Norse as complex as both, but a language somewhat less complex than both, as the Vikings acquired a robust but abbreviated version of English grammar.

#### 4.2. *The Pidgin Fallacy*

Our assumption, then, is that significant structural reduction can play as central a role in language contact as structural transformation or elaboration.

However, there is neither theoretical nor historical justification for an assumption that this reduction began with pidginization. This tendency appears repeatedly in language contact studies, and I suspect that this is because of the assumption that simplification is significant only in pidginization and creolization, while marginal or “controversial” otherwise. It follows, then, that if a contact situation has led to general simplification too anomalous to ignore, then pidginization must have played some part.

Given that the language under analysis is not remotely as simple as a pidgin today, a typical strategy is to assume that since its presumed pidginization, the language has expanded into a full language, that is, a creole, leading to analyses such as that of Bailey and Maroldt (1977), who proposed that Middle English was a creole. Another strategy is to suppose that after pidginization, prescriptive influence drew the language back toward the standard. An example here is Versteegh’s (1984) proposal that the difference between New and Classical Arabic is due to pidginization later reversed by influence from the Qur’an.

But absence of historical evidence leaves such accounts eternally speculative, and subject to observations of earlier inflectional erosion and absence of documented pidgin stages. In general, such accounts are founded upon a misimpression—that an intermediately simple language such as a creole can only have developed from the bottom up, starting as a pidgin. Creoles, as full languages, indeed represent a degree of simplification intermediate between the utter breakdown in pidgins and the typical baroque-ness of, for instance, Navajo. But because of the prominence of the literature demonstrating pidgins developing into creoles, such as the vast amount

of study of Tok Pisin and the high-profile controversy over Derek Bickerton's work in Hawaiian Creole English, there is a tacit sense that the pidgin-to-creole pathway is the only, principal, or "prototypical" way that creoles develop.

However, this is not true. Tok Pisin and its sister creoles indeed developed this way, just as Pidgin Sango in Africa is creolizing in similar fashion. I also believe (cf. McWhorter and Parkvall 2002) that this was the case for most plantation creoles, despite objections from scholars such as Michel DeGraff and Claire Lefebvre (cf. McWhorter 2004). But languages can become robustly simplified from the "top down" as well, without needing to touch base at a pidgin stage before rising again from the ashes. There is no evidence that Réunionnais French, African American Vernacular English, or Popular Brazilian Portuguese developed from pidgin precursors, for example. Rather, grammars can be simplified to an intermediate degree *from the outset*, with a preliminary dip into the pidginization stage playing no part in the process.

My interest in this book is in presenting a case that this grammar simplification is true, such that like arguments for particular languages will no longer be received as controversial and will occasion no confusion. For example, if Polome (1968) argues that the Shaba Swahili vernacular developed from a pidgin, and then Kapanga (1993) shows that Shaba Swahili is not simpler than standard Swahili to justify such an analysis, I seek that when De Rooij (1997) proposes that Shaba Swahili in fact is the result of widespread second-language acquisition of the standard that this is received not as a novel observation but as ordinary.

Erstwhile pidgins, therefore, are irrelevant to these intermediate cases. The argumentation will make only very occasional reference to Tok Pisin, Russenorsk, Foreigner Talk, the biopidgin, or other touchstones of pidgin and creole studies.

## 5. Assumption Three: Inflection Is but One Aspect of Grammatical Complexity

The reader is finally asked to avoid a possible tendency to assume that I am using "complexity" as essentially a shorthand term for inflectional morphology. Because most linguists are Indo-European speakers, and inflection has received so much academic attention, it is natural for inflectional paradigms to be the first thing to come to mind when complexity is mentioned. But inflections are but one of many aspects of elaborification addressed in this book. For example, in the English chapter, the features discussed include directional adverbs, V2, pronominal inventory, external possessor constructions, inherent reflexive constructions, and derivational morphemes; grammatical gender inflection is but one feature among these.

In fact, in terms of how complexity and language contact interact, inflection is a mere first layer. Afrikaans has lost most Proto-Germanic inflection but otherwise retains almost as much of the Proto-Germanic legacy as other Germanic languages. Creoles like Papiamentu and Haitian retain much of their lexifiers' derivation but have lost their inflections, while no creole retains lexifier inflections but not its derivation.

Because inflection is but one of many features important to the presentation, we err in supposing that the complexity referred to in this book can be applied to a famous complexity argued mainly for inflection, the cyclicity concept. It is well known that periphrastic constructions often become inflectional ones, which are in turn replaced by new periphrastic ones that in turn become inflectional, and so on. Thus Latin's inflectional future-tense markers evolved from Proto-Indo-European "to be" forms, which affixed to the root in what began as a periphrastic construction indicating the future (*\*kanta b<sup>h</sup>umos* [sing we-be] "We will sing"). Then in Vulgar Latin, a similar periphrastic form arose alongside the inflectional one with the *have* verb, *cantare habemus* "We will sing." The *have* forms then affixed to the verb to create French's *chanterons* (or Italian's *canteremo*, etc.), but in French a new periphrastic construction with another heavily used verb, "go," competes with the inflectional form (Hopper and Traugott 1993, 9–10):

TABLE 1.3. The Inflection Cycle from Proto-Indo-European to French

Pre-Latin		Latin		French
<i>*kanta b<sup>h</sup>umos</i>	>	<i>cantabimus</i>		
		<i>cantare habemus</i>	>	<i>chanterons</i>
				<i>nous allons chanter</i>

Presumably, in a hypothetical future stage of French, today's suffixal form would be replaced by a prefixal one such that the first-person plural prefix would be something along the lines of [zł̃], or perhaps given that *nous allons* is replaced in colloquial French by the third-person singular *on va*, a prefix like [vā], where the *on* ([ō]) segment eroded and left its nasality on the vowel of the following syllable. Hodge (1970) is a classic illustration of similar processes in various families.

However, for one, cyclicity accounts do not demonstrate that grammars go through stages in which they lack inflection entirely. They show the cycle as occurring with particular forms, such that we assume that at any given time in the grammar, various inflections are at differing points in the "cycle." As Dahl (2004, 279) remarks: "It is not always clear whether scholars who say things like 'linguistic evolution is cyclic, not linear' intend them to apply to specific phenomena or to languages as a whole." On the possibility that such linguists do intend that this cyclicity applies to the entire typology of a grammar at a time, Dahl goes on to point out that "it seems far-fetched to assume that languages become analytic because speakers are fed up with syntheticity and vice versa" (280).

And, in general, whatever our analysis of cyclicity with inflection, this framework does not apply to the many other aspects of complexity, often not even susceptible to the cyclic framework. For example, no one claims that grammars go through a "cycle" in which a highly elaborate pronominal inventory with dual, first-person plural exclusive, and honorific forms "erodes" into one marking three basic persons and not even number, and then builds back up into a more elaborate one. There is

no V2 “cycle.” A grammar that subdivides demonstratives between four or five gradations of distance (e.g., here, there, over there, way over there) is not expected to collapse these into a two-way proximal/distal distinction and then gradually build this back up into a five-way one. Overall level of complexity in grammars does not “cycle,” as the tour of the world’s languages made clear.

But what, after all, is “complex” when it comes to human language grammars? The reader will justifiably seek an outline of just what my conception of grammatical complexity entails. This is the subject of the next chapter.

# Defining Grammatical Complexity

Defining grammatical complexity in human language grammars is challenging. There is always the concern that designating one grammar as less complex than another may be based on incomplete data or superficial analysis, the latter possibly founded in a confusion between complexity and mere difference.

These are genuine problems, but I take issue with an implication in many assessments of the complexity issue that such problems render an extended address of the topic futile. This book will proceed upon a metric of grammatical complexity that, while hardly perfect, attempts to evade as much as possible the very real dangers in the enterprise.

Under my framework, grammatical complexity is measured according to three factors.

## 1. Three Aspects of Complexity

### 1.1. *Overspecification*

Languages differ in the degree to which they overtly and obligatorily mark semantic distinctions. I will term this difference one of *overspecification*.

For example, the Northern California American Indian language Karok has grammaticalized different verbal suffixes for various containment mediums: *pa:θ-kírih* “throw into fire,” *pa:θ-kúrih* “throw into water,” *pa:θ-rúprih* “throw in through a solid” (Bright 1957, 98, 102). (These morphemes are not perceivable reflexes of the words for *fire*, *water*, or *solid* respectively.) Most of the world’s grammars do not happen to have grammaticalized such fine-grained overt expressions of containment mediums, and it would be impossible to argue that Universal Grammar specifies such. On the contrary, as useful as these suffixes are in Karok grammar, their emergence was due to a chance elaboration within a particular semantic area, not communicative necessity. Thus, in the area of marking containment mediums, Karok is *overspecified* in comparison to

English, just as in its grammaticalized marking of definiteness and indefiniteness of NPs, English is overspecified in comparison to Karok, which lacks this trait.

No grammar is devoid of overspecification. That is, all grammars mark a number of distinctions that are not necessary to communication. For example, to the extent that most grammars mark a classic distinction between three persons in singular and plural in pronominals, inflection, and so on, they are overspecified, given that some languages mark only three persons with no specification for number (such as ones in Papua New Guinea [Dol 1999] and Indonesia). Thus any sense that the six-way distinction is “ground zero” is an illusion.

However, the claim in this book is that while all grammars are overspecified to some degree, all grammars are not equally overspecified overall. In other work, I have directly challenged the idea that a preliminary sense that Haitian Creole is less generally overspecified than French is an illusion, and that sustained analysis would reveal Haitian and English as equally overspecified. In this book, I challenge the related idea that all older languages are equally overspecified, that is, that a cursory mention of English’s fine-grained future marking strategies is enough to conclude that Karok and English balance out in terms of overspecification. I will show that grammars differ in degree of overspecification, and while small differences in degree are attributable to chance, where the difference is more extreme, this is due to the tendency for non-native acquisition to shave away features less necessary to communication, especially when this acquisition is difficult for adults to acquire.

Overspecification under this metric refers to a wide range of grammatical features encountered worldwide. I will present a nonexhaustive but representative list of the kinds of features that I intend, as a preparation for the argumentation that the five case studies is founded upon.

*Noun class marking.* Some grammars obligatorily mark nominals according to real-world categories, based on features such as animacy and shape (with such marking typically extending to concord with modifiers). Bantu languages are a classic example, such as in Swahili:

TABLE 2.1. Six Principal Noun Classes in Swahili

singular		plural		
I	m-tu	II	wa-tu	‘person’
III	m-ti	IV	mi-ti	‘tree’
V	ji-cho	VI	ma-cho	‘eye’
VII	ki-tu	VIII	vi-tu	‘thing’
IX	n-dege	X	n-dege	‘bird’
XI	u-moja	XII	u-moja	‘unity’

Such a grammar is overspecified in comparison to a language with no such marking. This trait does not require inflectional typology; numeral classifiers in isolating languages of Southeast Asia are analogous. However, many other grammars do not mark noun classes with morphemes free or bound (I am writing in one).

*Person/number marking.* Some grammars mark a finer grain of distinctions of person and number than others. In contrast to Indo-European languages, the Oceanic language Kwaio marks the dual and paucal, as well as an inclusivity distinction in the first-person plural (Keesing 1985, 28):

TABLE 2.2. Pronouns in Kwaio

	singular	dual	paucal	plural
first person (inclusive)	(i)nau	(‘i)da’a	(‘i)dauru	gia
first person (exclusive)		(‘e)me’e	(‘e)meeru	(‘i)mani
second person	(i)oo	(‘o)mo’o	(‘o)mooru	(‘a)miu
third person	ngai(a)	(‘i)ga’a	(‘i)gauru	gila

*Obviative marking.* There are languages that mark a “fourth person” in clauses with two third-person referents, obligatorily indicating which of the two is more prominent in the discourse. The Indo-European speaker might well marvel that a grammar would choose to mark such a distinction overtly, and indeed obviative marking is an interesting but unnecessary feature; an overspecification. Here is an example from Eastern Ojibwa (Rhodes 1990, 107):

- (1) Maaba dash oshkinawe o-gii-bawaad-am-n wii-bi-ayaa-**ini**-d  
 this EMPH young.man 3-PAST-dream-3INAN-OBV FUT-coming-be.at-OBV-3  
 myagi-nishnaabe-**an** x-wii-bi-nis-igo-waa-d-**in**.  
 foreign-people-OBV REL-FUT-coming-kill-INV-3-OBV  
 ‘Then this young man dreamed that foreigners would come to kill them.’

*Inalienable possession.* In many grammars, possessive marking differs for that which belongs to one in an inherent, eternal sense, as opposed to that which is formally and less permanently owned, as in Mandinka *i faamaa* “your father” versus *i la koloŋo* “your well” (Lück and Henderson 1993, 23). This distinction is hardly necessary to even nuanced communication, as is readily apparent to English speakers. It is an overspecification.

*Definiteness marking.* The highest estimate known to me of the percentage of the world’s languages that have grammaticalized articles (free or bound) at all is roughly half (Plank and Moravcsik 1996, 204), and Moravcsik (1969, 87, 93–98) estimates that only 39 percent of that subset have both definite and indefinite markers, and thus roughly one in five of the world’s languages have both markers. In this, then, the presence of these markers in English—as well as creoles, which tend to have definite and indefinite articles or at least one of the two—is an overspecification.

*Tense* Languages differ in the degree to which they subdivide time with overt marking. Kikongo, for example, marks four grades of pastness including the completive (Welmers 1973, 350), while Japanese has only one overt marker of past tense, and no grammaticalized indicator of completeness exclusively:



TABLE 2.3. Past Marking in Kikongo and Japanese

Kikongo	English	Japanese
<i>Nsuumbidingí nkóombo.</i>	'I bought a goat (today).'	<i>Yagi o katta.</i>
<i>Yásuumbidi nkóombo.</i>	'I bought a goat (yesterday).'	<i>Yagi o katta.</i>
<i>Yasáumba nkóombo.</i>	'I bought a goat (earlier).'	<i>Yagi o katta.</i>
<i>Nsuumbidi nkóombo.</i>	'I have bought a goat.'	<i>Yagi o katta.</i> (goat ACC bought)

Kikongo is more overspecified in this regard than Japanese — while Japanese is more overspecified than the few languages with no marking of tense whatsoever, such as Maybrat in Papua New Guinea (Dol 1999).

*Aspect.* Languages also differ in the degree to which they mark aspectual distinctions. English, for example, marks the habitual with the bare verb, contrasting with the progressive construction which marks present tense (*I watch television* signifies that the watching is a regular occurrence; only *I am watching television* signifies that the event is occurring at the moment). In this, English is more overspecified in habitual marking than its Germanic relatives as well as Romance, in which the bare verb marks both the present and the habitual, with interpretation typically left to context.

*Mood.* The subjunctive marking so familiar in European languages is, in the cross-linguistic sense, an overspecification. All languages have some mechanism for marking the hypothetical if explicitness is necessary. But for all subordinate verbs expressing hypothetical actions to require that the irrealis nature of the proposition be marked (French *afin qu'il le fasse* 'so that he does it') is just one choice a grammar might make, not a universal. Similarly, a language is overspecified compared to another one to the extent that it has distinct markers for the imperative mood (including negators exclusively used in the imperative) as opposed to using just bare verbs.

*Evidential marking.* Many languages require that the source of information expressed by a proposition be marked. Here are examples in the Amazonian language Tuyuca (J. Barnes 1984):

- (2) Kiti-gí      tii-      gí      'He is chopping trees' (I hear him)  
     chop.trees-he AUX      tí      'He was chopping trees' (I heard him)  
                                 í      'He is chopping trees' (I see him)  
                                 hɔi      'Apparently he is chopping trees' (I can't tell)  
                                 yigí      'They say he chopped trees'

These markers vary allomorphically, with distinct forms used for conversational participants, two genders of third-person singular, and plural, and then for present versus past tense. A grammar with evidential marking is overspecified in comparison to one where such distinctions are only indicated optionally.

*Copulas.* In addressing overspecification in regard to the copula, it is necessary to treat a constrained definition of the feature. Formally, the copula is a linker between subject and nonverbal predicate, as in *I am your father* and *I am in the house*. All

uses of the *be*-verb in a language, such as with progressive, perfect, and participial constructions, are not, formally, copulas. And then, cross-linguistically, languages recruit as true copulas not only *be*-verbs but demonstratives, focus markers, clitics, and even prepositions.

The very marking of the relationship between subject and nonverbal predicate is an overspecification, as demonstrated by the many languages like Russian in which they are absent in neutral sentences (*Ivan ø moj brat* ‘Ivan is my brother’). English is overspecified in marking the distinction with *be*, but then only does so invariantly according to the semantics of the predicate. Things are quite otherwise in many languages, in which there are different copula morphemes according to whether the predicate is equative or locative, such as Fongbe of Togo and Benin:

- (3) Ûn nyí Àfiáví.  
1SG be Afiavi  
‘I am Afiavi.’ (Lefebvre and Brousseau 2001, 144)
- (4) Wémâ ɔ ɖɔ távò jí.  
book DEF be.at table on  
‘The book is on the table.’ (ibid., 147)

Then other languages make even finer grades of copular distinctions. In the Cariban language Panare, copula marking differs according to distance from the speaker as well as animacy (Gildea 1993, 54–55):

- (5) maestro **kěj** e’ñapa  
teacher COP.ANIM:PROX Panare  
‘The Panare (here) is a teacher.’
- (6) maestro **něj** e’ñapa  
teacher COP.ANIM:DIST Panare  
‘The Panare (there) is a teacher.’
- (7) e’chipen **měñ** manko  
fruit COP.INAN mango  
‘Mango is a fruit.’

*Negators.* Grammars may develop a great number of monomorphemic negator allomorphs indexed to semantic distinctions. For example, the southern Min variety of Chinese has nine (Chappell 2001a, 346):

TABLE 2.4. Negator Allomorphs in Southern Min

<i>bô</i>	perfective
<i>m̃</i>	imperfective
( <i>íá</i> ) <i>bē</i>	negation of expectation
<i>boē</i>	negation of ability
<i>boài</i>	negation of perfective desiderative (‘didn’t want to V’)
<i>m̃mài</i>	negation of imperfective desiderative (‘didn’t want to V’)
<i>mài</i>	negative imperative
<i>m̃mó</i>	negative hortative
<i>m̃bién</i>	negation of necessity

*Demonstrative gradation.* The two-way proximal/distal distinction between *this* and *that* in English is underspecified compared to finer distinctions in many other languages. For example, the Austronesian language *Tukang Besi* of Sulawesi marks three grades of distinction, and then has separate forms in all three grades according to referentiality and the presentative (i.e., when the object is immediately presented) (Donohue 1999, 137):

TABLE 2.5. Demonstratives in *Tukang Besi*

	actual	referential	presentative
this	ana	meana'e	kaana'e
that	atu	meatu'e	kaatu'e
yonder	iso	measo'e	kaaso'e

*Ergativity.* Many languages must overtly mark that a subject is an agent rather than playing a less active role in the proposition, restricting one marker to agents (in transitive sentences) while marking nonagentive subjects (in intransitive sentences) with the same marker (or zero-marker) that patients receive. The Caucasian language *Andi* is an example:

- (8) (a) Voc:u-di      homologyi      vuq'o.  
         brother-ERG comrade.ABS brought  
         'The brother brought his pal.'
- (b) Voc:i          vuq'o.  
         brother.ABS came  
         'The brother arrived.' (Cercvazhe 1965, 226, cited in Harris and  
         Campbell 1995, 240)

In marking this distinction between types of subject, ergative languages are overspecified compared to languages that leave this distinction unmarked. Similarly, to mark experiencer subjects with a dative (or related) marker is an overspecification, such as in the Caucasian language *Tsez*:

- (9) año-r              meši      b-esu-s.  
         shepherd-LAT calf:ABS AGR-find-EVID  
         'The shepherd found the calf.' (Comrie et al. 2000, 19)

*Valence adjustment.* Languages differ in the degree of overt machinery that they devote to marking discourse prominence of subject over object. Passive constructions are one example of this impulse, but cross-linguistically, we encounter constructions that mark other shades of valence marking that languages like English leave to context. For example, inverse marking serves to show that a referent which would likely be subordinate under neutral semantics in fact played a stronger role. In this Cree sentence, the inverse marker determines that the meaning is not that the man killed the louse:

- (10) Ta:pwe: mac-a:yi:siyiniwe:sah nipah-ik o:hi ihkw-ah.  
 truly bad-person kill-INV this-OBV louse-OBV  
 'Truly the louse killed the evil man.' (T. Payne 1997, 212)

Meanwhile, antipassive marking demotes a patient's prominence via oblique case marking (counter to the literal semantics of that case) as well as other alternations, as in Yupik Eskimo (Payne 1997, 219):

- (11) (a) Yero-m keme-q nerre-llru-a.  
 Yero-ERG meat-ABS eat-PAST-3SG  
 'Yero ate meat.'  
 (b) Yero-q kemer-meng nerre-llru-u-q.  
 Yero-ABS meat-INST eat-PAST-INTR-3SG  
 'Yero ate (meat).'

Where the fact that the subject ate meat specifically is less important, *meat* is marked with an oblique case, here the instrumental.

*Focus marking.* Grammars differ in the degree to which marking of focus is obligatory. All utterances entail focus upon some constituent, but while some languages only mark this overtly when the focus is especially vehement (*It was the weather that ruined the event*), some grammars require, in overspecification fashion, indication of the focused constituent in all sentences.

The classic example is languages of the Philippines like Tagalog, which obligatorily marks focus on a sentential constituent with both a trigger particle *ang* and affixal markers that vary according to the constituent in focus. Here is the process illustrated in Tagalog (Schachter 1987, 941) (AT = agent trigger, PT = patient trigger, DT = dative trigger, BT = benefactive trigger, TG = trigger marker, AC = actor, DR = directional):

- (12) actor:  
 (a) **Mag**-aalis ng tindero **ang** bigas sa sako para sa babae.  
 AT.PROG-take out TG storekeeper PT rice DR sack BEN woman  
 'The storekeeper will take some rice out of a/the sack for a/the woman.'
- patient:  
 (b) Aalisi-**n** ng tindero **ang** bigassa sako para sa babae.  
 PROG.take out-PT AC storekeeper TG rice DR sack BEN woman  
 'A/the storekeeper will take the rice out of a/the sack for a/the woman.'
- directional:  
 (c) Aalis-**an** ng tindero ng bigas **ang** sako para sa babae.  
 PROG.take out-DT AC storekeeper PT rice TG sack BEN woman  
 'A/the storekeeper will take some rice out of the sack for a/the woman.'
- benefactive:  
 (d) **Ipag**-aalis ng tindero ng bigas sa sako **ang** babae.  
 PROG.take out-DT AC storekeeper PT rice TG sack TG woman  
 'A/the storekeeper will take some rice out of the sack for a/the woman.'

Thus sentences are not allowed that do not indicate which constituent is in focus, as opposed to languages where such distinctions are rendered optionally with intonation or are left to context.

*Pragmatic particles.* Grammars also vary in the degree to which they give overt marking to pragmatic implications. Languages of Southeast Asia are particularly rich in such particles. The Tibeto-Burman language Lahu, for example, has many such particles, which are central to basic expression, conventionalized into highly particular and idiosyncratic subdivisions of semantic and pragmatic space, reminiscent in both proliferation and in particularity of meaning to those in German like *schon*, *auch*, and *eben*. Here is one sentence from Matisoff (1973, 181) exemplifying the use of just some of the particles, with the translations provided by myself according to the grammar and checked with the author:

- (13) kólɔ̌ cɛ                      tí tɛ̌ qǒ ɔ̌, te mâ pè̌ tù hɛ̌.  
 Thai to-the-extent-that just EMPH as-for TOP do NEG finish FUT probably

‘If it’s really only the Thai (who are doing it), they’ll probably never get it finished!’

It may be useful to note that this is a thoroughly ordinary sentence of Lahu, not an unusually “expressive” one, and more important, there is no attempt in the glossing to “exotify” the sentence. For example, *cɛ* does not in any sense translate simply as “if,” but in its other uses clearly makes a semantic contribution leading Matisoff to term it an “extensive” particle (170–71); *tɛ̌* is not merely a focus particle, but one with the specific pragmatic purpose of averring the truth of a proposition in forceful observation or argument in a “See, there you go!” sense (171); and so on.

For a grammar to wend into overt and regular marking of these shades of attitude is an overspecification; real-life context known to conversational participants makes this degree of specificity unnecessary.

*Relativization.* For a language to have a distinct strategy for relativization, rather than using the same strategy for both relativization and subordination, is an overspecification. Thus English is overspecified compared to many languages with its distinction between *The man who did it* and *I saw that it wasn’t working*. The Algonquian language Montagnais, like many Native American languages, has no such distinction, for example, applying a conjunct verb form to both relativization and subordination (TA = transitive animate):

- (14) (a) (Ne) na:peu e-tshi:tu:t-et                      n-ukumish an.  
           that man        CONJ-leave-CONJ<sub>3S</sub> POSS-uncle that  
           ‘The man who is leaving is my uncle.’ (S. Clarke 1982, 140)
- (b) Ni-tshissen-im-a:u e-tshi:tu:t-et.  
       1S-know-TA-3S        CONJ-leave-CONJ<sub>3S</sub>  
       ‘I know he is leaving.’ (ibid., 138)

### 1.2. Structural Elaboration

An aspect of one grammar may differ from that aspect in another's in terms of the number of rules (in phonology and syntax) or foundational elements (in terms of phonemic inventory) required to generate surface forms.

*Morphophonemics.* For example, all natural languages have morphophonemic processes. But in Celtic languages, these entail a long list of consonant mutations, triggered by a wide range of grammatically central interfaces, in which synchronically, they are based not on assimilatory influence from preceding segments such as voicing (*leaf, leaves*), nasalization, and so forth, but are phonetically quite unpredictable. For example, in Welsh, *cath* "cat" occurs in citation form with *eu* "their," but with other possessive pronouns undergoes particular mutations:

- (15) eu *cath* 'their cat'  
 fy *nghath* 'my cat'  
 ei *gath* 'his cat'  
 ei *chath* 'her cat'

Notice that the mutation alone carries the functional load of distinguishing gender in *gath* versus *chath*. Thus the mutations often straddle a line between phonology and morphology, with its triggering on objects a similar case:

- (16) (a) Gwelodd ci.  
           saw           dog  
           'The dog saw.'  
 (b) Gwelodd Alun/ef/ø gi.  
           saw           Alun he he dog  
           'Alun/he saw a dog.' (Ball and Müller 1992, 195–96)

Few would claim that it was inappropriate to state that Welsh morphophonemics are more complex than English's. Welsh has more morphophonemic alternations, and they are conditioned by more interfaces in the grammar than in English. Moreover, they are usually phonetically unpredictable, requiring rote storage. Their existence is readily perceivable even to the untutored Welsh speaker, whereas for English speakers, the very existence of morphophonemics requires tutelage, and even then is often intuitively parseable as merely "making things easier to pronounce," which a Welsh speaker could not even venture to do of alternations like *fy nghath*. There are, simply, more and deeper morphophonemic processes between the phonemic and the phonetic in Welsh than in English. This is one of many features in which grammars can differ in terms of structural elaboration.

Even in more conventional cases of morphophonemics, grammars differ in terms of how many areas of grammar the processes apply to and their specificity. For example, in Kwaio, there are various genitive constructions marking assorted shades of relationship. In just the one marking detached parts of an object, the vocalic genitive

marker changes morphophonemically according to the final vowel of the referent. This means that if the part is *not* detached, such as in referring to a pig's belly, then the morphophonemic process does not occur:

- (17) oga-na boo  
 pig-GEN belly  
 'a pig's belly'

But if the part is detached, then the following alternations occur (Keesing 1985, 15–16):

- (18) oga oga-*e* boo 'a belly section of pork'  
 fote fote-*e* 'shoulder blade of'  
 lasi lasi-*i* 'head of'  
 lodo lode-*e* 'seed of'  
 nunu nunu-*i* 'shade of'

Note that with *o*-final nominals, alteration extends to the stem as well, with the final *o* changing to *e*. But then, this is suspended in the specific case that the preceding nominal is deverbal: *bulo* ("twist"): *bulo-e firi* "rope of domestic tobacco." Grammars differ in how much listing the morphophonemic aspect requires.

*Phonemic inventory.* A phonemic inventory is more elaborated to the extent that it has more marked members. Markedness is here intended strictly in reference to cross-linguistic distribution: marked phonemes are those encountered less frequently in the world's languages than others conventionally deemed unmarked; for example, ejectives, clicks, and labialized consonants versus stops, rounded back vowels, and glides.

The motivation for treating such sounds as metrics of complexity is not a claim that these are more articulatorily complex in the sense explored by Trubetskoy (1931); determining the precise measurement of articulatory complexity has proved too controversial to serve adequately as a foundation for this book. Rather, the motivation for using these sounds as a metric is implicational, in the Greenbergian sense: an inventory with a great many marked sounds (e.g., a click language) is more complex than one with all or almost all unmarked sounds (e.g., a Polynesian language) because *the former type of inventory has marked members in addition to unmarked ones*—the marked sounds *imply* the concurrent existence of unmarked ones (there exist no phonemic inventories with only cross-linguistically marked phonemes) (Greenberg 1966a, 1966b). A larger phonemic inventory requires maintaining finer intersegmental distinctions within the bounds of the human vocal apparatus. There is obviously no claim that marked phonemic inventories are more difficult to acquire natively; our reference is solely to the number of distinctions entailed by such inventories in comparison to others.

Here, then, is the phonemic (not phonetic) consonantal inventory of Tsez (Comrie, Polinsky, and Rajabov 2000):

Table 2.6. Tsez Phonemes

	stops			affricates		fricatives		nasal	liquid	semiv.
	v'less nonej.	v'less ej.	voiced	v'less nonej.	v'less ej.	v'less fric.	voiced fric.			
bilabial	p	p'	b					m		w
dental	t	t'	d	c	c'	s	z	n	r	
lateral				ɕ	ɕ'	ʃ			l	
palatal				č	č'	ʂ	ž			y
velar	k	k'	g							
uvular		q'		q		x	ɣ			
phar'ized	q <sup>ʕ</sup>	q <sup>ʕ</sup>					ɣ <sup>ʕ</sup>			
pharyng.						ħ	ʕ			
glottal						h				

Note that Tsez has plenty of common sounds; what is distinctive is the marked ones in addition: ejectives, uvulars, pharyngeals, pharyngealized uvulars, lateral affricates, and so on. Few will disagree that this is a more elaborated articulatory system than that of a Polynesian language with a mere dozen or so segments.<sup>1</sup> Nor is there any logical reason to suppose that qualitative difference in phonemic inventories must be binary: Tsez and a language like Hawaiian obviously constitute extremes. Overall, grammars differ gradiently on this measure of complexity, as on many others.

Similarly, a tonal system is more structurally elaborated than another one when it has more tones, because this phonology requires mastery and processing of a larger set of contrasts and requires maintaining finer intertonal distinctions within the bounds of the human vocal apparatus.

*Grammatical gender/declension and conjugational classes.* A grammar is structurally elaborated to the extent that nominals vary in their concordial requirements and/or case markers according to phonological traits of the root. The complexity is moderate in a language like Spanish, in which paradigms of case markers are absent, and only concord is of issue. The feminine, requiring articles *la* and *una* and adjectival suffix *-a*, is marked by various final segments or sequences such as *-a*, *-ion*, and *-tud*, and nominals are otherwise usually masculine. Yet this is obviously a complexity compared to English, in which grammatical gender is expressed neither with affixes nor free morphemes and does not exist at all.

Then, the elaboration increases in languages like Latin, in which concord was determined by a three-way gender distinction between masculine, feminine, and neuter, while case marking varied according to five declension classes distinguished by phonetic traits. There is no question of Latin's nominal morphology being merely one way of many of encoding a distinction expressed in a language like English by other means. Latin quite simply had a battery of machinery here for which there is no equivalent for English nouns.

Meanwhile, grammars may also subdivide verbs into conjugational classes, among which person/number suffixes vary. Thus French *parler*, *tu parles*, *vous parlez*, but *finir*, *tu finis*, *vous finissez* (\**finez*).



*Ergativity.* Ergativity increases not only a grammar's overspecification but its structural elaboration, in that ergativity typically shares space with nominative/accusative alignment, occurring only in, for example, the third person or the past. Thus a split-ergative grammar entails two strategies of morphological overspecification rather than one.

*Heterogenous word order.* Where a syntax entails regular flouting of its basic order of constituents for reasons grammatical rather than pragmatic, this tends to constitute structural elaboration.

V2 requirements in Germanic languages are an example. In German, basic order in matrix clauses is SVO: *Johann trank das Bier* "John drinks the beer." But the V2 requirement holds the verb in second place even when it is preceded by an element other than the subject: *Gestern trank Johann das Bier* "Yesterday John drank the beer." This is a rule in addition to the SVO one. Then the trait in some Germanic languages of leaving the verb at the end of subordinate clauses is another complication: *Ich sah, dass Johann das Bier trank* "I saw that John drank the beer."

Generative accounts stipulating that the subordinate verb remains clause-final because subordinator *dass* blocks its movement to COMP or I are more descriptive than explanatory, as I discuss in chapter 4. A question all such accounts leave is why it is so rare cross-linguistically that a filled COMP node leaves a verb stranded at the end of a sentence. For reasons genuinely unclear to me, much of the generative work on this issue seems founded in an assumption that the Germanic and Romance languages are the only ones in the world. If the Germanic COMP node is indeed filled by overt complementizers in such a way that a verb has nowhere to move to—a configuration presumably delineable via positing that the complementizer occupies particular branches of I—then we might treat this as a cross-linguistically peculiar configuration. This could be one of many possible manifestations of UG, under legions of which, it is clear, an overt complementizer does *not* leave verbs in subordinate clauses sentence-final. As such, it would qualify as an elaboration.

Clitics also often complexify word order, such as Romance pronominal object clitics, which "climb" into preverbal position despite the languages' SVO order (as opposed to equivalent object clitics in many other languages that remain postverbal). French *Je le sais* "I know it" contrasts with many languages in which the object clitic does not occur before the verb, as in Buginese *mu-ita-i* "you see him" (Sirk 1979, 18).

Another example is the tendency of clitics in some languages to seek a particular position in a clause regardless of their grammatical role. Serbo-Croatian clitics occur after the first accented constituent, and can even occur after the first accented word, intervening within a constituent. This placement requirement contrasts with the grammar's generally free word order:

- (19) Taj **mi**     **je**     pesnik napisao knjigu.  
       that to.me AUX poet wrote book  
       'That poet wrote me a book.' (Spencer 1991, 355)

A final example would be subject-verb inversion in interrogative sentences in many European languages, a trait quite rare outside of Europe (Ulan 1978) (this

would subsume *do*-support in English). Typically, languages indicate the interrogative with intonation and/or the appendage of an interrogative particle (Indonesian *apa dia sudah makan* [QU she PERF eat] “Has she eaten?” [Sneddon 1996, 311]).

### 1.3. Irregularity

Finally, grammars differ in the degree to which they are festooned with irregularity and suppletion.

*Grammatical gender and noun classes.* For example, grammatical gender marking entails not only complexification but irregularity that must be learned by rote. In Spanish, there are nominals whose phonology indicates one gender despite their concord being with the opposite gender: *la mano blanca* “the white hand,” *Buenos dias*, and adjectives invariant for gender, such as those ending in *-e* (*el hombre grande*, *la mujer grande* “the big man, the big woman”).

Then in German, gender is much less predictable than in Spanish; there are rules of thumb, but only that, such that essentially gender must be learned with each noun. The suffix *-ant* usually indicates the masculine (*der Komödiant* “the comedian”) but there are exceptions like *das Restaurant*. The suffix *-tum* is usually neuter (*das Eigentum* “the property”) but there are ample exceptions like *der Irrtum* “the error.” The student of German learns to treat *-e* as a feminine ending (*die Lampe*) but then must set aside cases like *der Friede* “the peace” and *der Name* “the name.”

Similarly, in languages like Bantu with extensive paradigms of noun class markers, the semantically predictable cases constitute overspecification, but the semantically unpredictable ones constitute irregularity. Such cases are quite common in such languages, such as Swahili *mdudu* “insect” in the noun class for humans, or the many nouns in the *u*-class, which usually contains abstract concepts (*umoja* “unity”), that denote collections of small things (*ugali* “porridge”). Meanwhile, there are morphophonological irregularities. The *n*-class prefix occurs only before *d*, *g*, *j*, and *z*, assimilates in some cases (*mbu* “mosquito”), but is absent in most (*asali* “honey,” *tende* “date”)—except that the prefix does occur (although possibly with assimilation) on all monosyllabic nouns (*nta* “wax”). Also, the *n*-prefix is invariable for number—albeit its corresponding subject concord prefix upon verbs is not invariable, and this prefix then varies according to animacy: *ilizi* for inanimates, *a/wa* (the prefixes for the human classes I and II) for animates.

*Plural marking.* Languages are also elaborified to the extent that plural marking includes irregularities. English’s closed set of irregular plurals like *children* and *people* is paralleled by Russian’s larger one, including examples like *syn* “son,” *synov’ja* “sons” (expected *syny*) and *doč’* “daughter,” *dočeri* “daughters” (expected *doči*).

But plural marking can be much more irregular in a grammar. In German, a masculine noun may take *-e* for the plural, but almost as equally will also take an umlaut: *der Arm*, *die Arme*, *der Besuch* (visit), *die Besuche* but *der Arzt* (doctor), *die Ärzte*, *der Gast* (guest), *die Gäste*. Similarly, a “regular” neuter is *das Jahr* (year), *die Jahre*, but then *das Buch*, *die Bücher*, *das Volk*, *die Völker* and even *das Bett* (bed), *die Betten*, *das Floss* (fin), *die Flösse*. There is no equivalent to this degree of irregu-

larity in pluralization of nominals in grammars that lack the inflection that occasions such disruptions.

Then there are languages where there are no perceptible rules for plural marking at all. Arabic “broken plurals” are an example. Pedagogical grammars regularly flag this as an indisputably learner-unfriendly feature, in capitulatory fashion after having frequently asserted beforehand that Arabic is not as difficult as it may seem. There is a regular suffixed plural in Arabic, but it is secondary; core nominals are marked for plural according to a long list of patterns of interaction between the trilateral root and surrounding vowels. Which pattern a noun will take is unpredictable, and must be learned by rote like grammatical gender in French and German. Here are just the most frequent patterns:

TABLE 2.7. Broken Plural Patterns in Arabic

singular	plural	
qalam	‘aqlām	‘pen’
bayt	buyūt	‘house’
kalb	kilāb	‘dog’
kitāb	kutub	‘book’
dawla	duwal	‘country’
šahr	‘ašhur	‘month’
wazīr	wuzarā‘	‘minister’
šadīq	‘ašdiqā‘	‘friend’

*Irregular root forms.* It is typical for grammars to include roots that alter unpredictably under certain morphological conditions. Indo-European speakers are especially familiar with irregular verbs: French *je bois* “I drink,” *vous buvez* “you (pl.) drink,” *je buvais* “I was drinking,” *je boirai* “I will drink,” *j’avais bu* “I had drunk.” Germanic strong verbs are especially illustrative, with certain verbs indicating the past with ablaut based on no predictable phonetic qualities, requiring a significant degree of rote acquisition: English *blink*, *blinked*, then *drink*, *drank* (a common [sink, sank] but not regular ablaut pattern), and then thoroughly suppletive examples like *think*, *thought*, with other past-marked forms phonetically akin to *thought* having present forms quite unlike *think* phonetically: *bought*: *buy*; *fought*: *fight*; *sought*: *seek*.

But the blot of such verbs is hardly a universal in grammars, especially ones in which boundedness does not give entrée to the ravages that phonetic juxtaposition threatens. In less inflected languages, irregularity tends to infect, at the least, a small class of heavily used verbs, but does not necessarily extend to dozens. For example, in Saramaccan Creole, the verb “to be” *da* is exceptional in occurring only in the present, with alternate copula *dé* occurring if the *be*-verb takes tense, aspect, or mood, and there is a portmanteau form when *gó* “to go” is marked with the progressive proclitic *tá*, *nángó*. But overall, verbal irregularity is but a whisper in the grammar.

*Suppletion.* Finally, grammars differ in the degree to which they exhibit suppletion. Suppletion is moderate in English, especially evident in the verb “to be,” which distributes various Old English roots across person, number, and tense: *am, are, is, was, were, been, be*. But the Caucasian language Lezgian has no fewer than sixteen suppletive verb forms (Veselinova 2003).

Elsewhere, suppletion is present in various corners of grammars. In French, combinations of genitive/associative *de* with an inanimate pronominal object are expressed with the suppletive portmanteau *en*: *J’en prends deux* “I take two **of them**” (\**Je prends deux de eux*), and similar combinations with *à* “to” require *y* (*Vous allez à la maison* “You go to the house,” *Vous y allez* “You go to it”). (Note also the structural elaboration, in that the resultant clitics must occur before rather than after the verb.) In Spanish, the third-person indirect pronominal clitic *le*, when preceding an object clitic, transforms to *se*: *Le dí un libro* “I gave a book **to him**,” but *Se lo dí* “I gave **it to him**” (\**Le lo dí*).

More important, throughout this book I will use the terms *complexity*, *overspecification*, *structural elaboration*, and *irregularity* based on their description above. That is, I will use *complexity* as a generic term covering the other three and only in this fashion. *Structural elaboration* will connote only the characterization I have attempted above, and will not be used as a mere alternative to “complexity,” of which I treat structural elaboration as one of three manifestations.

Thus at all times, where the reader may be moved to ask “What does he mean by complexity?” or “What does he mean by structurally elaborated?” my attempt at an answer is precisely in the preceding section of this chapter. At no point will I use these terms in heedlessly generic fashion, whatever the shortcomings of my definitions.

## 2. Inflection and Complexity

While inflection is but one aspect of complexity as I treat it in this book, it will loom large in all five case studies. This is because it happens to embody so vividly all three of the types of complexity that I will adduce, and is also especially well covered in even briefer grammatical treatments. As such, it will be useful to expand somewhat upon my above observations that inflections cannot be viewed as mere bound alternatives to morphemes elsewhere manifested as free forms, and as such, are irrelevant to assessing complexity.

To be sure, the expression of a distinction with an inflection is not more complex than its expression as a free morpheme in all cases. This is especially true of agglutinative inflection as opposed to fusional. For example, “I wanted” is *ni-li-taka* (I-PAST-want) in Swahili but *mi bi ké!* (I PAST want) in Saramaccan. That in Swahili the pronominal and the past morpheme are incorporated as a segmental sequences within a single word does not qualify as “complex,” in that they remain readily perceptible as discrete units of meaning, and are incorporated into words with little or no phonetic or paradigmatic disruption (morphophonemics and suppletivity are rather scattered and shallow in Swahili).

But inflections often mark distinctions unknown in any form in other grammars, such as noun class markers in Swahili. Then especially as we move toward the functional portion of the typological cline, inflections often vary allomorphically in interaction with phonology in a fashion with no equivalent in an isolating grammar, and tend strongly to devolve into irregularity that further increases their complexity quotient in comparison to an isolating language.

2.1. *Two Extremes*

The question we must be clear on, then, is the following: Is the sense that inflections tend to complexify a grammar based on a Eurocentric fetishization of boundedness as a mark of sophistication? Or, to put another spin on it, is the attention that I will pay to inflection due to an Anglophone’s exoticization of the declensional and conjugational paradigms so sparse in his native tongue?

Questions such as these must be asked. They must also be answered, and thus it will be useful to compare a grammar with no inflections at all, Saramaccan Creole, with a heavily inflected language, Estonian. This will allow the cleanest possible examination of the issue.

Estonian, like its more famous sister language Finnish, has a vast number of case endings on nouns and adjectives. Here is *raamat* “book” with the case endings in the singular:

TABLE 2.8. Cases in Estonian

nominative	raamat ‘book’
genitive	raamat-u ‘of the book’
partitive	raamat-u-t ‘some book’
illative	raamatu-sse ‘into the book’
inessive	raamatu-s ‘in the book’
elative	raamatu-st ‘out of the book’
allative	raamatu-le ‘onto the book’
adessive	raamatu-l ‘on the book’
ablative	raamatu-lt ‘off from the book’
translative	raamatu-ks ‘like a book’
terminative	raamatu-ni ‘as far as the book’
essive	raamatu-na ‘as a book’
abessive	raamatu-ta ‘without the book’
comitative	raamatu-ga ‘with the book’

Other than the wrinkle that the case endings beyond the genitive are appended to the genitive form (*raamatu*) rather than the nominative, one might treat these endings as mere equivalents of free morphemes in analytic languages. Thus *raamatu-ga* “with a book” would be equivalent to *ku búku* in Saramaccan, and so on. Under this analysis, *-ga* is not more complex than *ku* anymore than Swahili’s past inflection *-li-* is more complex than Saramaccan’s *bi*.

But this analysis would miss a great deal of complication, which renders highly questionable a case that Estonian’s inflections are irrelevant to assessing its overall

complexity in comparison to Saramaccan. Specifically, we will examine just one aspect of Estonian grammar, the genitive and partitive cases (data and description from Tuldava 1994 and Viitso 2003).

## 2.2. The Estonian Genitive

In the genitive, the Estonian alternation between nominative *raamat* and genitive *raamatu* is by no means a mere inflectional variant of Saramaccan's analytic *búku* versus *u dí búku* "of the foot." In fact, the formation of the genitive in Estonian is based on general tendencies rather than rules; generally the genitive form must simply be learned by rote like gender in most Indo-European languages.

Nouns and adjectives ending in a consonant may take, unpredictably, *a*, *e*, *i*, or *u* as the genitive ending:

ilusa 'pretty'  
ukse 'door'  
pliiatsi 'pencil'  
suitsu 'smoke'

Usually, roots ending in vowels do not change in the genitive:

isa, isa 'father'  
vana, vana 'old'  
õpetaja, õpetaja 'teacher'

But there are a great many exceptions, and not just in a few nouns of extremely high usage as in nouns with irregular plurals in English like *men*, *women*, and *children*:

nimi, nime 'name'  
meri, mere 'sea'  
veri, vere 'blood'

Then, most words ending in *-ne* take an *-se* genitive ending:

inimene, inimese 'person'  
naine, naise 'woman'

But then there are exceptions here where there is no change:

kõne, kõne 'speech'  
hoone, hoone 'building'  
laine, laine 'wave'

Meanwhile, morphophonemic consonant gradations, famous in Baltic-Finnic, complexify the situation even further. The gradations apply to second consonants as resyllabified upon suffixation. They are of two types, ones where consonants weaken and ones where they strengthen. Double consonants in the nominative weaken in the genitive via shortening:

lipp, lipu 'flag'  
pikk, pika 'long'

Stops also weaken in the genitive through shortening, by voicing or other changes in manner of articulation:<sup>2</sup>

pilt, pildi 'picture'  
 selg, selja 'back'  
 sada, saja 'hundred'

But when the root syllable is heavy or ends in certain consonant clusters, the final consonant does not just weaken but is absent:

kuub, kuue 'suit jacket'  
 laud, laua 'table'  
 koht, koha 'place'

However, when the consonant cluster ends with *d*, then there is either assimilation or (in the case of [ld]) elision of the second consonant:

vend, venna 'brother'  
 keeld, keelu 'prohibition'

And then note that even in these strong-to-weak gradated cases, the vowel that the genitive is expressed with remains unpredictable (there is no rule deciding that the genitive of *lipp* "flag" is *lipu* rather than, for example, *lipa*).

Then there are outright irregularities, of which this is but a very brief sample:

poeg, poja 'son'  
 tuba, toa 'room'  
 viga, vea 'error'  
 mees, mehe 'man'  
 süda, südame 'heart'

And these are just the cases where the nominative is strong. When the nominative has a weak second consonant, then it becomes strong in the genitive, according to a general rule that the nominative and the genitive have opposing gradations:

pale, palge 'cheek'  
 hammas, hamba 'tooth'  
 kääne, käände 'case form'

### 2.3. *The Estonian Partitive*

Then Estonian has its partitive case. Again, where the nominative form of "water" is *vesi* and the partitive form is *vett*, we might suppose that *Joo vett* "Drink some water" is a mere matter of *-tt* as a bound equivalent for Saramaccan's *só* in *Bebé só wáta!* Semantically, yes, but the affixation of the partitive suffix in Estonian is a much more complex affair than the mere utterance of invariant *só* is in Saramaccan.

The partitive suffix takes many forms. After a nominative root ending in a long vowel or diphthong, it is *d*:

maa, maad 'earth, ground'  
 pea, pead 'head'

But then, with a few words, when the diphthong ends with *u*, there is no change:

au, au ‘honor’  
nõu, nõu ‘advice’

If the nominative ends with a consonant, then often (but by no means always) the partitive is expressed with *t* added to the genitive form; here, the forms are nominative, genitive, and partitive:

tänav, tänav, tänavat ‘street’  
lennuk, lennuki, lennukit ‘airplane’

But often, if the genitive form ends in *-e*, this is dropped in the partitive:

õis, õie, õit ‘blossom’  
uus, uue, uut ‘new’

Then with a large but closed class of nominative roots ending in *n*, *l*, *r*, or *s*, the *t* is simply appended regardless of the genitive form:

keel, keelt ‘language’  
joon, joont ‘line’  
tütar, tütart ‘daughter’  
mees, meest ‘man’

But there are exceptions where *t* is appended regardless of the genitive form even to roots that do not end in *n*, *l*, *r*, or *s*, such as *õlu*, *õlut* “beer” (genitive *õlle*).

Meanwhile, many disyllabic words ending in *a*, *i*, or *u* are invariant across the nominative, genitive, and partitive:

ema, ema, ema ‘mother’  
kivi, kivi, kivi ‘rock’  
vaba, vaba, vaba ‘free’  
muna, muna, muna ‘egg’

While some disyllabic words ending in *a*, *i*, or *u* are invariant in just the genitive and partitive:

nimi, nime, nime ‘name’  
suvi, suve, suve ‘summer’

In some cases, the partitive form is the genitive one with a triple vowel (conventionally indicated in pedagogic Estonian orthography with ‘):

loom, looma, ‘looma [loooma] ‘animal’  
asi, asja, ‘asja [asssja] ‘thing’

Again, the weak/strong gradation difference adds further complications. The general tendency is that while the genitive tends to the opposite gradation from the nominative, the partitive has the same gradation as the nominative, thus “restoring” the weakening that occurs in the genitive form or, if the nominative form was weak and the genitive strong, then recapitulating the weak form of the nominative.



Below are some examples where the nominative strong form is restored—although in these cases affected by gradation there is an additional rule, that the partitive form has the same vocalic ending as in the genitive:

pikk, pika, **pikka** ‘long’  
jõgi, jõe, jõe ‘river’

But if the genitive form is irregular, then the partitive form is simply identical to the nominative regardless of the genitive form’s final vowel:

tuba, toa, tuba ‘room’  
nägu, näo, nägu ‘face’  
iga, ea, iga ‘age’

Then, when the nominative is weak, the partitive reproduces that weak form—but appending *-t*, this time with the genitive ending irrelevant:

hammas, hamba, hammast ‘tooth’

Meanwhile, a subset of disyllabic roots ending in *-si* take *-tt* (such that the *Joo vett!* cited above is, in fact, this type of exceptional form):

käsi, **kätt** ‘hand’  
vesi, **vett** ‘water’  
mesi, **mett** ‘honey’

Then there are some thoroughly eccentric exceptions such as:

sõber, sõbra, sõpra ‘friend’

#### *2.4. Inflections in Estonian and Saramaccan according to the Three Complexity Metrics*

Within the metric that this book is based on, the data above confirm that, leaving aside temporarily whether Saramaccan “balances out” elsewhere, Estonian’s genitive and partitive inflection is more overspecified, structurally elaborated, and irregular than Saramaccan’s.

##### 2.4.1. OVERSPECIFICATION

For example, Estonian does not apply genitive and partitive marking in mere equivalent fashion to Saramaccan’s distinction between *u dí búku* “of the book” and *só búku* “some books.” Rather, Estonian has grammaticalized its genitive and partitive marking into a greater degree of semantic overspecification than Saramaccan.

With singular objects, Estonian only uses the nominative in the positive imperative and with impersonals:

- (20) Võta raamat!  
take.IMP book.NOM  
‘Take the book!’

Only when the object is plural is it expressed with the nominative in neutral affirmative sentences (ILL = illative case):

- (21) Isa viis lapse-d kool-i.  
 father take.PAST child-NOMPL school-ILL  
 'Father took the children to school.'

If the object is singular in such sentences, then it is marked with the genitive:

- (22) Ma võtsin raamat-u.  
 I take.PAST book-GEN  
 'I took the book.'

Saramaccan does not extend genitive marking to mark objects:

- (23) \*Mi bi téi u dí búku.  
 I PAST take GEN DEF book  
 'I took the book.'

Semantically, we could see this as a mere difference in the semantic or syntactic domain of genitive marking. There is no reason to suppose that extending genitive marking to marking objects (given that Estonian has no accusative marker) is "more complex" than restricting the genitive to the possessive and attributive. However, Estonian's restriction of the genitive marking to the singular is a structural elaboration compared to the neutrality of genitive marking to number in Saramaccan.

And then, in negative sentences, Estonian uses neither the nominative nor the genitive but the partitive:

- (24) Ma ei võt-nud raamat-ut.  
 I NEG take-PART book-PART  
 'I did not take the book.'

In Saramaccan, partitive marking in the equivalent sentence could only encode a particular, literally partitive meaning:

- (25) Mi á bi téi só búku.  
 I NEG PAST take some book  
 'I did not take some/a few books.'

Nor has Saramaccan grammaticalized the partitive into the negative; it is used also in affirmative sentences:

- (26) Mi bi téi só búku.  
 I PAST take some book  
 'I took some/a few books.'

Thus Estonian has a three-way split in case marking according to semantics that is absent in Saramaccan, in which the objects in all three of the above sentences have no case marking at all:

- (27) (a) Téi dí búku!  
 take DEF book  
 'Take the book!'

- (b) Mi bi téi dí búku.  
I PAST take DEF book  
'I took the book.'
- (c) Mi á bi téi dí búku.  
I NEG PAST take DEF book  
'I did not take the book.'

Importantly, it must be noted that other uses to which Estonian puts this distinction indeed parallel similar distinctions that Saramaccan marks with different machinery. In Estonian, genitive-marked objects connote completion while partitive-marked ones connote the imperfective:

- (28) (a) Ma luge-si-n raamat-u läbi.  
I read-PAST-S book-GEN through  
'I read the book (all the way) through.'
- (b) Ma luge-si-n raamat-ut.  
I read-PAST-S book-PART  
'I was reading the book.'

But then, Saramaccan marks the same distinction overtly, not with case marking but with verbal particles, right down to *kabá* "finish" as the equivalent to Estonian's *läbi*:

- (29) (a) Mi bi lési búku kabá.  
I PASTread DEF book finish  
'I read the book through.'
- (b) Mi bi tá lési dí búku.  
I PAST IMP read DEF book  
'I was reading the book.'

This is a demonstration that a facile equation of morphology with complexity is indeed mistaken. The equivalence between (28) and (29) would not be immediately clear from brief perusals of grammars of Estonian and Saramaccan, because while the Estonian distinction in (28) would be demonstrated in one place because of its hinging on a morphological opposition, a Saramaccan grammar would present the complete and imperfective constructions as but two of several tense and aspect distinctions encoded with various free morphemes within VP. Claims of complexity differentials must be made very carefully.

But in this case, what remains is a basic fact: Estonian submits the object to more distinctions of semantic markings than Saramaccan, in a fashion that Saramaccan does not parallel in any sense anywhere in its grammar.

#### 2.4.2. STRUCTURAL ELABORATION

Estonian genitive and partitive marking is also more complexified than in Saramaccan.

The marking in Estonian is sensitive to phonology, phonotactics, and then also tied to morphophonemic variations of gradation. Saramaccan's free partitive morpheme *só!* is invariant, but genitive (*f*)*u* does vary morphophonemically with pronouns:

TABLE 2.9. Morphophonemic Variation of *fu* + Pronominal in Saramaccan

/fu mi/	>	[u mi]	'my'
/fu i/	>	[fi i]	'your'
/fu ẽ /	>	[fẽ ẽ]	'his/hers/its'
/fu u/	>	[fu u]	'our'
/fu unu/	>	[fu unu]	'your' (pl.)
/fu de/	>	[u de]	'their'

But otherwise, genitive (*f*)*u* occurs as *u* with full NPs (*dí táfa u Rohit* "Rohit's table"). These behaviors of (*f*)*u* certainly demonstrate that Saramaccan, like all natural languages, has complexity. But where genitive marking in Saramaccan entails morphophonemic alterations sensitive to six pronouns, Estonian genitive marking is sensitive to all nouns, pronouns, and adjectives, including distinctions based not only on the simple distinction between vowel versus consonant but also on more fine-grained distinctions of segmental quality, according to whether the segment (or cluster of segments) is weak or strong.

Then also, Estonian's division of labor between genitive and partitive occasions further complexities. The past negative requires the past participial form *lügenud* rather than only preposing *lugesin* with free morpheme *ei*. Certainly Saramaccan has a resultative participial form, encoded with reduplication:

- (30) DÍ bóto dé láilái.  
 DEF boat be.at load.load  
 'The boat is loaded.'

But this construction is used to express the resultative. It is in Estonian as well, but Estonian contrasts in extending it to usage with objects in negative sentences. Estonian, in applying its resultative to negative sentences in contrast with affirmative (but not imperative) ones applies the construction to an overspecification of the negative imperative—along with partitive marking—that is unknown in Saramaccan in any form.

Certainly there is no implication that Estonian's grammaticalization of the partitive in negative sentences is cross-linguistically unique. French, for example, has a similar distinction encoded with *de*: *Je lis des livres* "I read books" versus *Je ne lis pas de livres* "I don't read books." But for one, neither Saramaccan nor any creole conventionalize partitive marking to this extent; French's partitive marking, notoriously tricky to non-native acquirers given its dogged signaling of a distinction most languages restrict to more explicit contexts, is a typical symptom of an older grammar. Second, even French encodes the partitive with a morpheme that occurs either free or in portmanteaus in which it is phonetically perceptible (*des*), lacking the extensive allomorphy, specifications to phonological particularities of roots, and ample irregularities that Estonian's affixal expression of the same distinction conditions. That is, inflection is not always a mere substitute for free morphemes. It strongly tends to complexify grammars, and vastly so.

## 2.4.3. IRREGULARITY

Estonian genitive and partitive marking also occasions a great deal of irregularity that must be learned by rote. In Saramaccan, partitive marking is entirely regular, and genitive marking only occasions a shard of irregularity.

For example, when a possessive PP with *(f)u* in Saramaccan is a predicate, the copula is omitted, unlike with other PPs:

- (31) (a) Dí pindá ø u mí tatá.  
 DEF peanut GEN my father  
 'The peanut is my father's.'  
 (b) Dí wómi **dé** a dí wósu déndu.  
 DEF man COP LOC DEF house inside  
 'The man is in the house.'

But this quirk in one syntactic position hardly equals the rampant irregularities in Estonian's genitive marking.

Saramaccan also has one morphophonemic process where the connection between underlying and surface forms has become opaque, where the locative marker *a* combines with third-person singular pronominal *ê* to yield *nêê*. The relationship between *a* and *nêê* in the sentences below is hardly phonetically transparent:

- (32) A sindó **a** dí líba fêê.  
 he sit LOC DEF top POSS.it  
 'He sat on top of it.'  
 (33) Mi á wā baási mi nê sikiífi **nêê**.  
 I have a balloon my name write on.it  
 'I have a balloon with my name written on it.'

This contrast is the result of the fact that *a* is derived from an earlier *na*, as is clear in historical documents of Saramaccan:

1778:

- (34) dem no mussu komotto **na** Jerusalem  
 they NEG must leave LOC Jerusalem  
 'They must not leave Jerusalem.' (Schuchardt 1914, 2)

The initial consonant has since worn away from the item in isolation, but persists in the portmanteau morpheme. The result is a morphophonemic rule stipulating an [n] in combination with a particular pronominal form.

However, the rule has yet to become completely opaque. The *na* form is still used with deictic adverbs: *na akí* "on here," *naãndé* "there, on there," *na alá* "over there, (on something) over there" and some interrogative markers, such as *naúntê* [at-which-time] "when." Speakers also use *na* variably on its own in its isolated usage:

- (35) I sá njā **na** dí júu i ké.  
 you can eat LOC the hour you want  
 'You can eat at whatever time you want.'

A process like this shows signs of being a recent innovation not yet complete. But the irregularities in Estonian are thoroughly entrenched in all of their opacity, and overwhelmingly more numerous and wide ranging than the ones in Saramaccan, occasioning vast tables of morphophonemic rules indexed to phonetic segments, beside which exist innumerable exceptional cases. Such tables are unnecessary in a description of Saramaccan.

The above demonstration ought to temper any discomfort occasioned by a general tendency in this book to treat the spread of analyticity in a grammar as a sign of decreased complexity. Quite often, analyticity entails not simply the substitution of free morphemes for bound ones, but the unraveling of phonological and paradigmatic complexifications occasioned by boundedness.

### 3. Can One Grammar Be More Complex Than Another?

The above demonstrations also situate us for a preview of the type of reasoning that will lead in this book to assessments of one language as more complex than another overall, rather than in certain areas of grammar.

What we have seen shows, at the very least, that a claim that Estonian does not surpass Saramaccan in the complexity of its genitive and partitive marking would be hopeless. But this also leads us to a larger issue. The massive complexities of Estonian inflection do not stop at genitive and partitive marking, but pervade the grammar. Yet according to the idea that no grammar can be less complex than another, we must assume that Saramaccan has complexities on the order of Estonian elsewhere in its structure.

But the question is: where? Demonstrations that Saramaccan has, simply, complexity are unremarkable and irrelevant. The question is whether Saramaccan *equals* Estonian in complexity in other areas of its grammar. And here, we find that where Saramaccan exhibits complexities, Estonian has complexities of its own that “match” Saramaccan and leave the complexity gap suggested by the genitive and partitive area intact.

#### 3.1. Saramaccan Morphophonemics?

Relevant here, for example, would be an objection that Saramaccan has morphophonemic processes beyond the particularities with (f)u. And of course this is true: there exist no human languages without morphophonology, and it would be distinctly odd to encounter one where morphophonology applies only to a single preposition.

But our question, as always, is not whether Saramaccan is the simplest possible language. Since no such claim has been made, the simple fact that Saramaccan has morphophonemic processes is irrelevant to the question we are addressing. That question is whether Saramaccan’s *degree* of morphophonemic processes “balances out” with Estonian. It does not.

Thus in Saramaccan the fusion of third-person oblique pronominal clitic *ẽ* with preceding verbs that end in [a] or [ɛ] leads the preceding vowel to assimilate to it:

*paká* + *ê* “pay him” > *pakêê*, *túwε* + *ê* “throw it” > *túwêê*, but not with other preceding vowels:

*bebé ê* ‘drink it’  
*subí ê* ‘climb it’  
*sindó ê* ‘sit it down’  
*tjókó ê* ‘stab it’  
*lukú ê* ‘look at it’

This behavior is even sensitive to tone, in that it does not occur with the occasional verbs that have low-tone codas (*bà* + *ê* “gather it” > *bà ê*). The subject pronouns also usually fuse with the following negator in portmanteau forms: /mi á/ [I NEG] > [má]; /a á/ [he NEG] > [á]; /u á/ [we NEG] > [wá]; and so forth.

But these aspects of Saramaccan are local to particular interfaces between six pronominals. Such tendencies are quite common in any natural language with phonetically light items of heavy usage. These hardly constitute the same degree of complexity as genitive and partitive marking in Estonian, applying to nouns, pronouns, and adjectives as a class, and comprising phonetically distinct markers conditioned by fine-grained phonetic distinctions, these further complexified by aspects of gradation utterly unknown in Saramaccan, any other creole, or, in fact, most of the world’s languages.

### 3.2. *Saramaccan Tone Sandhi?*

So, if morphophonemics cannot make a case that Saramaccan is as complex as Estonian overall, what else in Saramaccan might help us? One place we might look is its tone sandhi, which has occasioned a fair amount of academic discussion (Ham 1999; Kramer 2002; Good 2003). Under an assumption such as Arends’s (2001) that creoles may harbor complexity in terms of the generation of rules rather than their surface manifestations, Saramaccan tone sandhi would seem to qualify as a useful candidate, as well as in that transcriptions of the language do not mark the sandhi, such that it could qualify as a complexity not readily apparent in most treatments.

But our question is whether Saramaccan tone sandhi is complex to an extent that parallels Estonian genitive and partitive marking. Upon examination, it does not.

Saramaccan high tone spreads rightward but breaks at the left edge of maximal projections (Ham 1999). Thus sandhi breaks before the prepositional phrase (# = sandhi blockage):

- (36) /dí sitónu tá náki mi a mí fútu/ → [dí sítónú tá náki mi # a mí fútu]  
 DEF stone IMP hit me LOC my foot  
 ‘The stone is hitting me on my foot.’ (Rountree 1972, 323)

By Ham’s analysis, assuming that subjects are generated as the left branch under maximal projection IP with the verb generated in the rightward branch as part of I’, sandhi spreads from subject to VP because there is no intervening maximal projection. On the other hand, while sandhi is not blocked between verbs and object

clitics like *mi*, because the clitics are contained within the VP, sandhi does break between verbs and full NP objects, these presenting a maximal projection:

- (37) /mi wási déé koósu/ → [mi wási # déé kóósu]  
 I wash DEF.PL clothes  
 'I wash the clothes.' (ibid.)

Nor is the system perfectly regular. For example, a few high-usage adverbs fail to block sandhi from a preceding verb, such as *hėti* 'not far':

- (38) /mi wáka hėti akí/ → [mi wáká hėti # akí]  
 I walk not-far here  
 'I walk right over here.' (ibid., 322)

There are also some words with inherent low tone impermeable to sandhi, such as past marker *bi*: [mi á bi lési], but \*[mi á bí lési].

One conclusion here might be to suppose that even if Saramaccan lacks the morphological elaboration of Estonian, its tone sandhi refutes a claim that Estonian is a more complex language. But this represents a rather abbreviated reasoning, which does not stand up to further reflection. The issue is one of degree.

And in that light, Saramaccan tone sandhi is essentially a regular, and even simple process, in which low tones become high like preceding ones—an unremarkable kind of assimilation—except where a maximal projection intervenes. The exceptions are scattered and few. Estonian genitive and partitive markers occur in several allomorphs, determined with fitful predictability according to multiple phonological specifications of which the weak/strong grade opposition is just one, and just as often, quite unpredictable (such as the vowel of the genitive ending as well as the completely exceptional cases). The volume of rote learning is massive.

### 3.3. Estonian Is More Complex Than Saramaccan

So, let us pull the camera back and assess equivalent sentences in Estonian and Saramaccan, with as inclusive a perspective on complexity as possible.

TABLE 2.10. Genitive and Partitive Marking in Equivalent Estonian and Saramaccan Sentences

Estonian	Saramaccan
Loe raamat läbi! read.IMP book. NOM through 'Read the book!'	Lési dí búku! read DEF book 'Read the book!'
Ma luge-si-n raamat-u läbi. I read-PAST-S book-GEN through 'I read the book.'	Mi bi lési dí búku. I PAST read DEF book 'I read the book.'
Ma ei luge-nud raamat-ut. I NEG read-PPbook-PART 'I did not read the book.'	Mi á bi lési dí búku. I NEG PAST read DEF book 'I did not read the book.'



Saramaccan differs from Estonian in the tone sandhi, breaking in these sentences between VP and object, and is inapplicable to past marker *bi* as one of the closed class of words with inherent low tone.

But Estonian's genitive and partitive marking remain, allocated specifically according to mood (indicative) and polarity (negative), and marked according to the intricate and exception-laden rules described above. For example, *raamat* is *raamat-ut* in the partitive only because it ends neither in a long vowel nor diphthong (which would condition a *d* ending), nor a short vowel (which would probably mean that the genitive and partitive markers were identical, although the vowel would be unpredictable), nor *l*, *n*, *r*, or *s* (which would likely mean that *t* was appended with no preceding vowel), nor *-si* (which would condition the ending *-tt*), and does not happen to be one of the words where the partitive is marked by tripling the length of the vowel, is not an outright exception, and does not happen to end in a consonant or consonant cluster subject to consonant gradations. Then, the fact that the vowel is *u* is unpredictable.

There are now several decades of literature describing Saramaccan, and nowhere in it is there any grammatical feature requiring this much description. In addition, Estonian distributes these sentences across three different verb forms, while Saramaccan uses but one.

To be maximally fair, Saramaccan has definite and indefinite articles, marking definiteness and referentiality obligatorily while Estonian does not. But Saramaccan expresses this with invariant free morphemes (*dí* "the," *wā* "a"). And then, Estonian "balances out" here, in that gradation affects verbs as much as nouns. The past form *lugesin* "I read" is based on the strong grade infinitive root *luge-*, but the present must occur in the contrasting grade, such that "I read" in the present is the weak grade *loen*. But then if the infinitive is weak, then the present is strong: *hakata* "to begin," *hakkan* "I begin." On top of this, the graded forms of these verbs are very often not predictable (*luge-* versus *lo-* for "to read"). Then Estonian has its irregular verbs, ranging from suppletivity (*minema* "to go," *lähen* "I go") to assorted less phonetically disruptive cases varying morphophonemically in a fashion only exhibited in verbs in particular tenses: regular *istuma* "sit," *istu-sin* "I sat," but the initial *s* of the past suffix is absent with disyllabic verbs with stem ending in *e*: *tulema* "to come," *tul-in* "I came" (*\*tule-sin*).

Indeed, there is irregularity in tense and aspect marking in Saramaccan verbs. But for our purposes, the crucial fact is that it is minimal. Identificational equative *da* "to be" becomes *dE!* when marked for tense or aspect:

- (39) (a) Mi **da**      dí      kabiténi.  
           I    COP   DEF   captain  
           'I am the captain.'
- (b) Mi **bi**      **dé**      dí      kabiténi.  
           I    PAST COP   the captain  
           'I was the captain.'

But this is a quirk of an allomorphic allocation (and between a mere two allomorphs), rather than a matter of an independent suppletive form: *dé* is elsewhere the copula used in locative, existential, and even class equative predications (*Mi dé wā*

*kabiténi* [I COP a captain] “I am a captain”). Then when the verb *gó* is marked with the imperfective, the aspect marker *tá* becomes *nán-* and binds to the verb: *mi nángó* “I am going,” as opposed to *Mi tá wáka* “I am walking.”

On top of this, Estonian’s marking of completive aspect is more grammaticalized than in Saramaccan. While Saramaccan distinguishes completivity with *kabá* “to finish” as in (29a) above, bare verbs can also carry completive meaning; *kabá* is used when explicitness is desired. Thus *Lési dí búku!* can mean either “Read the book through!” or “Read the book for a while!” and *Lési dí búku kabá* would mean roughly “Read the book all the way through (i.e., don’t get up till you finish it)!” However, in Estonian the distinction is, in most contexts, marked obligatorily when semantically applicable, with the *läbi* satellite. If the meaning is to simply spend some time reading the book, then the partitive is used: *Loe raamatut* “Read the book (for a while, with no need to finish it).”

### 3.4. Intuition versus Science

On the intuitive level of the interested Anglophone layman, Estonian grammar is a gorgeous but endlessly treacherous mess, appearing so complex and riddled with irregularity that one might wonder how humans could even speak it. Of course, speakers of related Baltic-Finnic languages would not have this impression, but then speakers even of highly inflected, morphophonemically rich languages beyond Baltic-Finnic would, since the complexities of Estonian are different from those they are familiar with (I have heard such an assessment from a Russian, for example). No one would find Saramaccan grammar intimidating in this fashion, despite occasional features presenting a challenge such as the tone sandhi. Nor is this a mere matter of inflection alone, given common tales of woe from speakers of Cantonese who encounter Mandarin, or from European tradesmen residing on the West African coast in the seventeenth and eighteenth centuries flummoxed by the tones in analytic languages like Twi.

My complexity metric can be seen as confirming that these gut-level impressions of Estonian and Saramaccan are based in reality, rather than being Eurocentric illusions. In terms of overspecification, Estonian is a language with tight linkage between form and semantics of the sort that elsewhere conditions features like ergativity with its distinction of types of subject. Saramaccan meanwhile leaves a great deal of such distinctions to context. Meanwhile, Estonian also obligatorily marks other distinctions largely left to context in Saramaccan, such as the partitive (or elsewhere, evidential hearsay marking). In terms of structural elaboration, Saramaccan morphophonemics seem scattered wrinkles, almost always phonetically intuitive as moderately disruptive assimilations. In contrast, Estonian morphophonemics are vastly transformative processes requiring a high degree of rote stipulation, muddying almost any sentence one wants to utter. Estonian word order is heterogenous in terms of verb placement; Saramaccan’s is typically SVO. In Estonian, irregularity is, essentially, a norm; in Saramaccan, it is exceptional.

That is, it is unreasonable to propose that Saramaccan has no complexity. However, it is reasonable to propose that Estonian is a more complex language than Saramaccan.

This book is predicated upon the idea that grammars vary in their degree of complexity according to the same kind of reasoning that leads to this conclusion about Estonian and Saramaccan. As such, I will argue that where complexity differentials are stark between one language and its relatives, non-native acquisition played a decisive role in its history, eliding the learner-unfriendly phenomena that pervade a language like Estonian.

## Epistemological Caveats

Before we proceed to the case studies, I would like to address three fundamental objections of a philosophical nature that my framework may elicit.

### 1. Linguistic Egalitarianism

Much possible resistance to this thesis will stem from a commitment in academic linguistics to countering the folk assumption that indigenous languages are “primitive.” In particular, my frequent reference to inflectional loss may appear to court some early linguists’ equation of inflections with evolutionary advance, such as Wilhelm von Humboldt’s view of the analytic drift in many Indo-European languages as a curious and regrettable decay. This linguistic egalitarianism, in Newmeyer’s (1986) terminology, among linguists was one of the inspirations of the very foundation of the Linguistics Society of America in 1924 (*ibid.*, 41), and was couched in a more general aim to dispel the notion that some cultures are less advanced than others, as pioneered by Franz Boas and his students such as Margaret Mead.

However, throughout this book I attempt to make clear that I am well aware that all human languages are highly complex. I am as dismayed as all other linguists at misimpressions that a language without inflections is a “simple” one, as in the case of someone I once met who had learned Twi in the Peace Corps and described it as having “no grammar” despite its complex tonal system. I find the same absurdity as any responsible linguist in the notion that a language without inflections, even if complex to some degree, must be less complex than one with inflections. I immediately think of languages like Min (Taiwanese) Chinese, replete with semantic overspecifications beyond the inflectional, and with each word carrying its own tone sandhi specifications.

Indeed, however, I will argue that while all languages are complex, all are not equally complex, that the differentials between them can be stark enough to merit attention, and that these differentials can be argued to trace to external intervention

rather than chance. I am not alone among linguists in questioning the orthodoxy that all languages are equally complex. Experienced field linguist and grammarian R. M. W. Dixon articulately presents my foundational assumption (1997, 75):

While it is the case that all languages are roughly equal (that is, no language is six times as complex as any other, and there are no primitive languages), it is by no means the case that they are exactly equal. I have done field work on languages in Australia, Oceania and Amazonia and they were certainly not all equally difficult to describe. There is no doubt that one language may have greater overall grammatical complexity.

As such, the difference I treat is not binary, but relative and clinal. I make no claim that any language is the simplest possible. This claim would, after all, be scientifically contradictory under my assumption that all languages accrete needless complexity over any period of time, and in any case, my proposition is that extensive non-native acquisition abbreviates complexity rather than eliminates it. Moreover, since the current state of the languages in the case studies is centuries or millennia removed from the periods when their transmission was interrupted, they have inevitably accreted new complexities in the interim.

Therefore, I would hope that the reader would receive this book as a statement targeted at a community of scholars within which the awareness that there exist no “primitive” languages is long confirmed and considered an unremarkable assumption. I am hoping that the reader will be open to the possibility that this assumption is not incompatible with a thesis referring to *gradient* complexity.

## 2. Incomplete Data?

Other objections to my approach might question whether I neglect certain crucial aspects of complexity.

### 2.1. Context as Complexity

One might argue, for example, that underspecification increases rather than decreases complexity, in creating ambiguity that must be resolved by the hearer. Hawkins (1985), for instance, notes differences between German and English occasioned by German’s tight linkage of case to semantics, as in *Dir gefällt das?* [to-you pleases that] “Do you like that?” as opposed to English usually not marking the experienterhood of subjects. Hawkins argues that German’s close association between form and semantics allowed by its rather rich case morphology is a kind of complexity, but that English has an equal complexity in requiring reliance upon context for parsing semantics in the absence of such morphology.

To the extent that one is inclined to treat reliance upon context as a type of complexity, one must view my conception of complexity as a subset of a larger one that includes the context factor. As such, one might treat this book as addressing less than the whole of complexity thusly conceived.

However, I maintain that non-native acquisition has a decisive and regular effect upon the concrete aspect of complexity as I have conceived it, and that this effect is an abbreviative one. It will remain urgent to incorporate this phenomenon into models of language contact even if we assume that the abbreviation merely yields to the alternative “complexity” of contextual recovery. It would seem that if we propose a module of complexity involving contextual recovery, then this is so qualitatively distinct from the concrete complexity I address that it has little bearing upon the coherence of my thesis.

Even if we assume that there are assorted interactions between the concrete and contextual aspects of complexity, these interactions no more invalidate restricting an argument to concrete complexity issues than the interactions between phonology and morphology invalidate a book treating phonology only. To wit, it is unclear that any of the arguments I will adduce could be shown to be flawed by virtue of neglecting particular aspects of contextual recovery.

And even then, we might ask just what evidence has been adduced that there are specific interactions between underspecification and cognitive awareness of the concept in question. No author has devoted a book, or even an article, to a presumed structured and complex relationship between grammatical form and the intricacies of contextual recovery. It would seem that contextual recovery would be essentially a matter of aspects of real-world cognition equal in all humans and always operating regardless of the nature of the grammar a person is using. For example, we assume that whether a grammar marks a two-way proximity distinction in demonstratives (*this/that*) or a finer-grained five-way proximity distinction, all speakers readily process the fact that a distal referent five yards away is closer to them than one twenty yards away.

Thus we assume that all humans have equivalent cognitive processes. But this means that to propose that contextual recovery is a kind of grammatical complexity entails believing that there is a mechanism besides general cognition, that forges a link between gaps in grammatical specification and their cognitive referents. But no authors present evidence to this effect, likely because there is no indication in linguistic or psychological data suggesting such machinery. Certainly we would expect that authors examining the Sapir-Whorf hypothesis, addressing how some grammars mark certain categories more obligatorily than others, would at least occasionally posit such specific pathways between underspecifications and the concepts they apply to. But they do not (cf. the absence of such proposals in Lucy’s [1992] summary of studies of the Sapir-Whorf hypothesis, or Gumperz and Levinson’s [1996] anthology). Rather, reports such as that from Davidoff, Davies, and Roberson (1999) showing that a language with discrete terms for two colors occasions a more ready distinction of the colors than a language with one term for both colors imply that the latter language leaves the speaker reliant upon general cognitive awareness assumed to be universal, not that this language entails an additional processing “patch” linking, for example, the “grue” term to the two concepts of green and blue.

And in any case, the contextual complexity concept would only apply to the overspecificational aspect of complexity under my model. For example, a conception such as Hawkins’s addresses mostly inflectional issues, whereas my model appeals to a much wider array of features. In terms of structural elaboration, for example,

reduction of morphophonemics and phonemic inventory does not require increased reliance on context. Because grammatical gender and declensional classes do not denote real-world distinctions, their eclipse does not require contextual recovery, given that the features never carried meaning at all. And the ironing out of heterogeneous word order can be argued to *decrease* reliance on context in favor of the regularity of grammatical roles occurring in a single order. Meanwhile, reduction of structural irregularity is irrelevant to context: for example, since French verbal irregularities in the *-ir* class such as the *-iss-* affix (*finissez*) mark no semantic distinction, their absence in the semicreole French of Réunionnais entails no contextual recovery.

## 2.2. *Pragmatics and Intonation*

The reader is correct to observe that my argumentation largely neglects pragmatics and makes no reference at all to intonation. I have not included these features for the mundane reason that they are insufficiently covered in most grammatical treatments for meaningful comparative analysis.

However, in response to a possible objection that pragmatic and intonational features might weaken my assessments of relative complexity between grammars, we must ask whether the inclusion of such features for a grammar I argue to be strangely less complex than its sisters' would render it complex *on a par* to them. For example, I argue that English is less complex, according to my metric, than any other Germanic language. This argument does not include pragmatics or intonation. But in surmising that the inclusion of these features would reveal English as equally complex to Icelandic, we must first consider that Icelandic, as well as all of the other Germanic languages, has its pragmatic and intonational modules as well.

Thus if pragmatics and intonation are just *equally* complex in English and its sisters, then properly addressing these features in all of the languages would leave the overall complexity differential I propose intact. This means that to suppose that my neglect of pragmatics and intonation refutes my arguments requires a particular assumption: that for some reason, a language with less complexity in the features that I do cover must *necessarily* have *more complex* pragmatics and intonation than its sisters, according to a presumed imperative to "balance out" with them in overall complexity.

But it is difficult to see the scientific basis of such an assumption. Why would English surpass its sisters in complexity precisely in pragmatics and intonation, but not in other features? What established and independently developed framework in linguistic science predicts, explicitly or epiphenomenally, that pragmatics and intonation must increase in complexity as underspecification, structural elaboration, and irregularity erode? Or, what linguistic principle rules out the existence of a grammar with moderate underspecification, structural elaboration, and irregularity and also only moderately complex pragmatics or intonation?

For any other features that one might identify as neglected under my analysis, the same questions apply. Given that the sister languages exhibit these features as well, whence the assumption that, for example, English must *surpass* them in complexity in just that feature? There is, at the end of the day, no scientific motivation for the idea that all languages are complex to exactly the same degree.

Thus the reader is correct to regret that for each language at the center of the case studies, I give little or no attention to pragmatics and none to intonation. But I submit that we must view this within an awareness that the language's sisters exhibit such traits just as vibrantly. We must at all times ask ourselves just why we might assume that the case study language must *surpass* its sisters in the complexity of these features.

I respect that some will feel nevertheless that an argument such as mine is incomplete until we have information on pragmatics and intonation in, for example, Frisian, Sudanese Arabic, Buginese, Hakka Chinese, or the many dialects of Kurdish. But the relevant data has not been gathered. Meanwhile, the pattern indicated by the data that is indeed available on the relevant languages is so inexorable that I feel it scientifically appropriate to apply them to a larger thesis.

I make no claim to address an ideally complete data set. But short of a logically rigorous argument as to why the inclusion of the features I neglect would necessarily refute my perspective, rather than bolster it or simply leave it intact, I submit that my thesis stands—albeit subject to refutation from other perspectives.

### 3. The Issue of Audience

#### 3.1. “*Real Linguistics*”

My argument will neither compel nor convince the reader who feels that substantial linguistic argument must be couched in the frameworks of generative phonology and syntax, and focus primarily upon analysis of the generation of rules in particular grammars.

This approach would, in fact, be antithetical to the goals of the argument. The generative paradigm is designed to demonstrate what all languages have in common; I am devoted to illuminating how languages differ despite their essential commonalities. Because of the massive diversity among languages, a paradigm seeking what they have in common will necessarily focus on a highly constrained collection of features, such as head movement, anaphoric relations, INFL, the lambda calculus, the faithfulness constraint, and so on. My paradigm, however, requires attention to features such as differentials in phonemic inventories, inalienable possessive marking, evidential marking, number of copulas, and demonstrative gradation. Such features are usually considered tangential to leading generative concerns, and, where they do figure in discussion, they are for reasons unrelated to those that center my argument.

Certainly there are instances in which generative positions may question my analysis of a given construction as overspecified, structurally elaborated, or irregular. For example, if differing verb movement patterns across Germanic are conditioned to a large extent by whether or not the COMP node happens to be filled, then what appears to be a complex array of differences on the surface presumably traces to a rather simple underlying alternation. Then, of course, many suppletive forms can be treated as “generated in the lexicon.”

But the question is whether such explanations could cover more than a small subset of the totality of my argumentation. The ideal refutation from “East Coast



linguistics” would be a demonstration that all of the differentials I treat are traceable to alternate parameter settings, syntactic configurations, or ordering constraints, and thus unamenable to gradation according to “complexity.” But it is difficult to conceive of how such an analysis could cover more than a sliver of the features I treat. Depth and specificity of morphophonemics and paradigmatic irregularity are features that would be especially resistant. Or, what more general aspect of structure could we analyze fine-grained gradation of demonstratives, or a proliferation of noun classes, as “falling out of”?

This brings us to an observation relevant to linguists beyond the generative realm as well: the interactional nature of linguistic modules and structures does not render impossible that grammars can be ranked in terms of complexity. To assume so is, in fact, a logical leap with no demonstrated basis in fact or theory. In reality, these very interactions are part of what is lost as complexity erodes; for example, morphophonemics bridging phonology and morphology, or ergative alignment bridging semantics and syntax. Rather, the only assumption that fact and theory predict is that such interactions *might* explain *some* of the contrasts between grammars that I will present. If these kinds of interactions instead explain *all*, or even *most*, of the features in my argument, then this is the unexpected case. As such, that case must be comprehensively demonstrated. That is, a passing suggestion or two does not constitute a refutation. This will especially be true if the suggestions are based on things such as the fact that phonetic erosion can complicate morphology—given that this is not always the case and is definitely not in the case studies I treat—or that contrasting verb movement patterns within Germanic can be traced to syntactic parameter settings (an argument fragile in broad view as applied to English, as argued in chapter 4).

This book, then, is concerned with qualitative differences between grammars, couched in an assumption that the universal traits manifested in parametrical alternations posited by the generative school are an interesting but small core, which each grammar decorates further with a vast amount of fortuitously emergent overgrowth.

One just might sense that my analysis of this overgrowth is inherently less “sophisticated” than generative paradigms. And indeed, tabulating that Persian has no grammatical genders while Pashto has four, or that Kerinci Malay has allomorphically variant noun pairs distributed according to sentential position in contrast to standard Indonesian’s invariant noun roots, is a less “abstract” endeavor than submitting disparate features of a single grammar to an analysis based on a posited “little *v*” node, or submitting aspects of Semitic nonconcatenational morphology to a hierarchy of constraints according to Optimality Theory.

But my thesis stands, quite simply, according to whether the difference between the features I compare can be traced regularly to extensive non-native acquisition. Therefore, even if my framework were so concrete and descriptive that a chimpanzee could assist me in my research, its validity would still be assessed by whether the adduced phenomena correlate significantly with sociohistory, period. When it comes to the particular interaction between sociohistory and structure that interests me, the generative paradigm, focusing on what grammars share, is for all its intricacies too crude for meaningful application beyond certain modules (such as inflectional complexity, as masterfully argued by Kusters [2003]).

Robinson (1975, 102–3) wrote of Noam Chomsky: “He has lost that wonder at the splendid multiplicity of language and languages which is the other side of the wondering at language as a common human possession.” I would not consider the absence of that “wonder” a “loss” in Chomsky’s followers; their wonder is simply focused on other things. However, Robinson captures a fundamental difference in interest between Chomskyan linguists and many others. To the typical Chomskyan, the differences between languages—intuitively fascinating to the amateur language buff—are a distracting illusion to argue away, via revealing underlying unities. The Chomskyan is at fault in this no more than is the scholar of pathways of energy production within cells who has little interest in the difference between the mating rituals of toads and sparrows.

However, I do question any corollary sense that, as it were, the student of toads’ and sparrows’ mating rituals is on a fool’s errand and would be making a richer contribution to science by joining the molecular biologist in exploring the production of energy, out of a sense that, in the end, this will predict and explain the mating rituals in any case. Certainly the mating rituals are driven by and founded upon the molecular processes—but just as certainly, the molecular processes alone will be of little use in explaining why the toad’s mating processes differ from the sparrow’s.

### 3.2. *Macrosociolinguistics versus Microsociolinguistics*

This book will similarly disappoint the reader interested in the details of how language structure differs according to sociological dynamics “on the ground.” There will be neither quantitative analysis of the Labovian sort nor analysis of issues such as the signaling of identity via linguistic variables. Rather, my interest is of a constrained sort: whether differences in complexity between related grammars can be attributed to extensive non-native acquisition. I assume that to the extent that my argument compels, it will be considered applicable more to theories of diachronic linguistics than sociolinguistics.

More precisely, the case studies address interactions that occurred too far back in antiquity for there to exist detailed accounts of the dynamics of interpersonal and cross-class interactions that typically interest the sociolinguist. My goal is to argue that differences between grammars can be responsibly traced to non-native acquisition even in ancient instances such as these, via comparison of the modern results combined with careful inferences from fragmentary but useful historical data.

In fact, I assume that transformations of the kind that I address occurred before widespread literacy and education retarded language change among ordinary people via the effects of prescriptivism. Modern English, Mandarin, colloquial Arabic, Persian, and Malay emerged among peoples largely illiterate or semilliterate, in societies in which the written word was a marginal luxury largely confined to commercial records, liturgical documents, and high folkloric literature. Crucially, at the relevant times, there existed no society in the world in which literacy was as widespread as it is in many societies today. My interest is in whether we can draw correlations between language structure and sociohistory in this period of human history despite the gaps in the data due to writing having yet to provide us with detailed accounts of sociolinguistic interaction.

Thus this is a book of detective work, on contexts for which only scattered concrete evidence of sociological factors remains for analysis. My address of non-native acquisition necessarily neglects, then, particulars of interactions between language and power, or the variable competences of individuals splitting the difference between the “code” of the “gatekeeper” and the lively vernacular of indigenous expression. Surely such sociological interfaces were rife in the relevant contact situations. But before the establishment of Western academic conceptions of social science, almost no one on earth was moved to record such observations. I posit that the tabulation of such things, while vital to sociolinguistic theory, is not necessary to the thesis of this book.

Specifically, I consider microsociolinguistic concerns to potentially refute my arguments solely to the extent that one might propose exactly how data on language and power or quantitative variation according to class, race, or gender would contest my claims. Otherwise, I assume that microsociolinguistic issues are a subsidiary factor in the contact situations I address, tantalizing but unrecoverable, entailing no challenge to the lines of my conclusions.

# English

## *The “Irritation” of Nuance*

### 1. Introduction

In the emergence of Modern English, simplification dominated complexification to a greater extent than in any other Germanic language. I will argue that the evidence suggests that this simplification was no a happenstance peculiarity, but due to a sociohistorical factor hindering the full transmission of the grammar across generations.

I will not revive the hypothesis of Bailey and Maroldt (1977) (followed by Domingue 1977; Poussa 1982) that Middle English was a creole that developed when Norman French invaders learned English imperfectly and expanded their reduced English into a full language. This hypothesis was motivated partly by the heavy admixture from other languages in English’s lexicon and derivational apparatus. But lexical mixture itself does not equate with creolization. Languages can borrow massive amounts of lexicon and even morphology without evidencing any traits that would suggest the label *creole* to any linguist, such as many languages of Australia (Heath 1981), and “mixed” or “intertwined” languages like Michif (Bakker 1997) and Media Lengua (Muysken 1997).

Bailey and Maroldt and their followers also based their argument on English’s notorious paucity of inflection in comparison to other Germanic languages. However, inflection alone, while a valuable sign of the larger process I will outline, cannot serve as the heart of the argument. Thomason and Kaufman (1988, 263–342), for example, note that inflectional loss had proceeded considerably before the Norman invasion, and even in dialects not in contact with French. In addition, today’s Mainland Scandinavian languages, unaffected by any contact as heavy as that caused by the Norman invasion in England, are only a little more inflected than English. Moreover, during the Norman rule, French speakers were but a numerically small elite, whose rendition of English can hardly have had impact on a vast majority of monolingual English speakers. These and other arguments will be taken as conclusive in this paper.

Thus inflection alone is susceptible to the idea that its heavy loss was due simply to English being fortuitously less “conservative” than its sisters, uniquely subject to a certain “drift.” Meanwhile, to the extent that English manifests other features less complex than equivalents in its sister languages, the generative historical linguistics tradition tends to ascribe these to “chain-style” effects of inflectional loss that modern syntactic theory would predict. Examples include the rich literature on the loss of OV and V2 word order, and other features such as obligatory postposing of particles to the verb (e.g., Platzack 1986; Van Kemenade and Vincent 1997; Fischer et al. 2000). The common consensus among specialists on the history of English is that features suggesting a break in transmission or unusual simplification in the English timeline are mere *trompe l’oeils*, having, in fact, emerged by ordinary processes of change.

However, the focus on certain abstract syntactic features that the generative enterprise conditions has perhaps narrowed our purview, in a fashion neglecting other aspects of grammar that suggest a larger story. I will propose that loss of inflection is but the tip of the iceberg in terms of Germanic features that English has shed, complemented by many other losses unconnected with analyticity. Overall, a comparison with its sisters reveals English to be significantly less complex, according to my metric. For example, English manifests less overspecification, such that Hawkins (1985) notes: “Where the surface structures (morphology and syntax) of English and German contrast... English surface structures exhibit less correspondence with their semantic representations than do those of German” (6), and later “German speakers are forced to make certain semantic distinctions which can regularly be left unspecified in English” (28).

Some scholars, such as Lass (1987, 317–32), recognize that English departs considerably from the Germanic template, but leave aside the question as to why, with the implication that the issue was a matter of chance. However, I will argue that a contact-based, external explanation provides a principled account for the relevant facts.

## 2. Complexity as a Matter of Degree

Importantly, my argument is not that English is in any sense *radically* decomplexified in a cross-linguistic sense. Its *do*-support, subtle subdivision of the semantic space of the future between four constructions (*I leave tomorrow*, *I’m going to leave tomorrow*, *I’m leaving tomorrow*, *I will leave tomorrow*), the subtle interplay of definiteness and referentiality underlying the use of its definite and indefinite articles, and other features are fatal to any argument that English is somehow a “simple” language.

Thus my argument proceeds in full acknowledgment of such features, but is predicated upon a *relative* argument: that overall, English is *less* complex than *its sisters*, to an extent suggesting something other than unbroken internal development.

My intention can be illustrated with an informal but heuristically useful observation. For both English speakers learning another Germanic language and speakers of other Germanic languages learning English, much of the acquisition task entails learning alternate, but in no sense more complex, strategies for expressing concepts.

An example would be the arbitrary differences in semantic space that prepositions cover: *pale with fear* versus German *blass vor Furcht*, and so on. However, it is my impression that for the English speaker, most of the acquisition task beyond this entails learning to attend to things English does not mark overtly, while for the speaker of another Germanic language, most of the task entails learning *not* to attend to features that their native language does mark overtly.

### 3. English and Its Sisters

#### 3.1. Inherent Reflexivity Marking

Germanic languages overtly mark what is often called “inherent reflexivity.” These differ from literal reflexives in that while these refer to an event involving two participants, in which both participants are involved, contrary to general expectation, the two participants are found to be the same entity (*He shot himself*), inherent reflexives entail a perception of one participant, performing upon itself an action whose reflexivity is the expected case rather than an anomaly (*He bathed*) (Haiman 1983; Kemmer 1993).

Inherent reflexives are the product of the grammaticalization and bleaching of the reflexive element in conjunction with verbs connoting inherently reflexive actions, such that in many languages, inherent reflexivity is marked in motion verbs (German *sich beeilen* “to hurry”), psych-verbs (*sich erinnern* “to remember”), and verbs of social behavior (*sich benehmen* “to behave”); this is a cross-linguistic developmental tendency (cf. Kemmer 1993; Peitsara 1997).

Inherent reflexivity marking is common in all of the Germanic languages but English, and was also present in early Germanic languages such as Gothic (*ni idreigo mik* “I do not repent” [Dal 1966, 155]) and Old Norse (where it was already grammaticalized to the point of morphologization; cf. 5.3), suggesting that it was a Proto-Germanic feature. In this and similar tables, for reasons of space, “Mainland Scandinavian” is represented by Swedish, the feature having also been identified in Danish and Norwegian.<sup>1</sup>

TABLE 4.1. Inherent Reflexives in Germanic

German	<i>sich rasieren</i> ‘to shave,’ <i>sich beeilen</i> ‘to hurry,’ <i>sich erinnern</i> ‘to remember’
Dutch	<i>zich scheren</i> ‘to shave,’ <i>zich bewegen</i> ‘to move,’ <i>zich herinneren</i> ‘to remember’ (203–4)
Frisian	<i>hy skeart him</i> ‘he shaves,’ <i>ik skamje my</i> ‘I am embarrassed,’ <i>ik stel my foar</i> ‘I imagine’ (66, 147)
Afrikaans	<i>hy bevind hom</i> ‘he is situated (at),’ <i>hy roer hom</i> ‘he gets going,’ <i>hy herinner hom</i> ‘he remembers’ (288–91)
Scandinavian	<i>raka sig</i> ‘to shave,’ <i>röra sig</i> ‘to move,’ <i>känna sig</i> ‘to feel’ (105–6)
Icelandic	<i>koma</i> ‘to come,’ <i>komast</i> ‘to get to, reach,’ <i>snúa sér</i> ‘to turn around,’ <i>skammast sín</i> ‘to be ashamed’ (105, 143)
Faroese	<i>raka sær</i> ‘to shave,’ <i>snúgva sær</i> ‘to turn,’ <i>ætla sær</i> ‘to intend’ (117–18)
Yiddish	<i>bukn zikh</i> ‘to bow,’ <i>shlaykhn zikh</i> ‘to sneak,’ <i>shemen zikh</i> ‘to be ashamed’ (89–90)

Note that a pronominal form restricted to reflexive use (e.g., a cognate of German *sich*) is not necessary to inherent reflexive marking. The absence of a reflex of *sich* is not local to English, but was already the case in Ingvaemonic, such that Frisian also lacks a *sich* reflex; Afrikaans does as well. In all Germanic languages, inherent reflexivity is marked in the first and second persons with the corresponding accusative or oblique pronoun (*ich rasiere mich, du rasierst dich* “I shave, you shave”), and in the third person, those without a *sich* reflex mark reflexivity with an accusative or oblique third-person pronoun, for example, Frisian *hy skeart him* “he shaves.”

In Old English, inherent reflexivity was marked with either the dative or the accusative pronoun in all persons:

- (1) þa **beseah** he **hine** to anum his manna and cwæð  
 then look.PAST he him.DAT to one.DAT his man-PL.GEN and say.PAST  
 ‘Then he looked at one of his men and said’ Visser 1963, 146)
- (2) **Reste** ðæt folc **hit** on ðam seofþan dæge  
 rest.PAST the people it.ACC on the.DAT seventh day  
 ‘The people rested on the seventh day’ (ibid., 147).

Even at this early date its use was optional ( $\emptyset$  where the reflexive pronoun would occur):

- (3) se sylfa Drihten wolde  $\emptyset$  of heofenum on eorðan beseon  
 he self Lord want.PAST from heaven.DAT on earth.ACC see.INF  
 ‘The Lord himself wanted to look upon the earth from heaven.’ (ibid., 146)

Throughout the Middle English period, however, inherent reflexivity was marked increasingly less (Mustanoja 1960, 431), likely preferred as a metrical device (Fischer 1992, 239). As early as Old English, texts suggest only vague semantic distinction between a given verb’s usage with and without the reflexive pronoun (B. Mitchell 1985, 114; Visser 1963, 322; Rissanen 1999, 256).

By the Early Modern English period, Peitsara (1997, 303) finds inherent reflexive marking in only a third of potential cases from 1500 to 1570, and in less than a sixth from 1570 to 1640. By the latter period, Peitsara finds the marking only in a limited number of verbs, including ones of motion, posture, self-care, and equipment; of psych-verbs only *fear* retains it (optionally) and among social ones, *commend*. Eventually, inherent reflexive marking with simple pronouns is eliminated completely, except for in scattered frozen archaisms (*Now I lay me down to sleep*).

As inherent reflexive marking declines in Middle English, *self*, which begins in Old English expressing emphatic reflexivity (among other uses: cf. B. Mitchell 1985, 115; Faltz 1985, 18–19, 35), increases in frequency; by the fifteenth century its use had bleached semantically into compatibility with verbs that previously took just a simple pronoun (Peitsara 1997, 320–23). Thus where Wycliffe in the fourteenth century has *Adam and his wijf hidden hem fro the face of the Lord God* (Old Testament, Genesis 3:8), Tyndale in the sixteenth century has *And Adam hyd hymselfe and his wyfe also from the face of the LORde God* (Five Books of Moses, Genesis 3:8). However, instead of extending to the full range of verbs that the bare-pronoun reflexive strategy once covered, *self*-pronouns settle into the modern pattern, largely marking only lit-

eral reflexivity. Today, reflexivity is usually marked where operation upon the self is emphasized for clarity (*bathe oneself*) or stylistic purpose (*they hid themselves*). With some verbs, the reflexive usage has conventionalized into a particular meaning (*to behave* versus *to behave oneself*), while in only a limited number of verbs is the marking obligatory (*pride oneself*, *perjure oneself*).

The result of this process was that as Peitsara neatly puts it, English became unique among Germanic languages in “an individual tendency to treat overt reflexivity as redundant, unless marked for practical or stylistic reasons” (1997, 337). This cannot be attributed simply to the fact that inherent reflexivity marking was already optional in Old English, given that such an account begs the question as to why English did not instead choose to conventionalize the initially optional usage rather than eliminate it. For example, the feature was also at first variable in German (Curme 1952, 155–56), but was eventually obligatorified; Curme (331) notes “German is usually tenacious of reflexive form even after its meaning has changed.”

Obviously this was also the case in the other Germanic languages. In Mainland Scandinavian, the grammaticalization went so far that the reflexive pronoun has eroded into a mere suffix on many verbs, creating deponents such as *minna-s* “to remember”; Icelandic and Faroese’s *-st* suffix is similar (and is also found in Nynorsk and other western Norwegian dialects; Peter Trudgill, February 2002, personal communication). Cornips (2002) documents the development of reflexive *zich* into a middle marker over the past hundred years in the Limburg dialect of Dutch (*Dit bed slaapt zich goed* “This bed sleeps well”). The English situation must also be seen within the context of a similar generalization of inherent reflexive marking across Europe as a whole. Haspelmath (1998, 276), for example, describes the development of anticausative marking with the reflexive pronoun as a pan-European Sprachbund feature. Our question, then, is why English took so anomalous a path as to eliminate the feature after having partially conventionalized it.

Along those lines, the grammaticalization of inherent reflexives is obviously connected to the marking not only of shades of reflexivity, but to distinctions of valence (transitivity) and mood (passive and middle voice). These distinctions are commonly related cross-linguistically (e.g., Lyons 1968, 373–75), and the Scandinavian *-s*-marked verbs, for example, also encode passivity (*bakas* “to be baked”). In this light, the disappearance in English of inherent reflexive marking can be seen as one symptom of the general drift toward “transitivization” that Visser (1963, 127–35) describes, where the overt distinction between transitive and intransitive use of verbs erodes. Under this analysis, the eclipse of inherent reflexive marking was part of a general process which also included the disappearance of the *ge-* prefix that once distinguished transitive verbs (*æman* “to run,” *geæman* “to reach, attain by running”).

While we might be tempted to suppose, as Visser’s treatment implies, that this “transitivization” drift was the fortuitous result of various independent processes that just happened upon a single result, the question we must ask is why similar processes did not converge upon the same result elsewhere in Germanic. Scandinavian also lost its *ge-* prefix, and the coalescence of vowels that once marked causative distinctions (Old English *sincan* versus *sencan*) is hardly unique to English. Yet if there is a tendency for a grammar to fill in “open spaces” in syntax as well as phonetic inven-



tories as Visser surmises (*ibid.*, 135), then we might ask why English did not, as its sisters did, seek to compensate for the ravages of phonetic erosion by obligatorifying its usage of reflexives to retain overt signaling of transitivity and passiveness. Put another way, why did English not submit these pronouns to “exaptational” usage, in Lass’s (1990) conception borrowed from evolutionary biology, recruiting reflexive pronouns as valence and mood markers as dozens of other languages were concurrently doing across Europe?

Instead, English became a grammar markedly less overspecified than its sisters in this area, leaving inherent reflexivity, transitivity, and causativity to context to an extent unique in its subfamily, and unusual in the Indo-European family as a whole.

### 3.2. *External Possessor Constructions*

When a possessed object falls into a semantic class roughly definable as inalienable, Germanic languages typically encode the possessor as an argument distinct from the possessed NP itself, as in German’s *Die Mutter wäscht dem Kind die Haare* “the mother washes the child’s hair.” While this construction is sometimes termed the “sympathetic dative,” I follow Vergnaud and Zubizarreta (1992) and König and Haspelmath (1997) in referring to this as the *external possessor construction*, as the term captures the larger generalization that the “external” constituent can be marked with cases other than the dative. “External possessive” generally refers to animate possessors, and applies to body parts, relatives, clothing, habitations, and sometimes even emotional conditions: the motivating factor is membership in the “personal sphere” (*ibid.*, 530–33).

This feature is found in all of the Germanic languages but English and Afrikaans; in Mainland Scandinavian and Icelandic the marking is locative rather than dative, while Yiddish has the dative for pronouns and the benefactive for nouns. Because Proto-Germanic surveys are notoriously sparse in their coverage of syntax, I have no data on this construction in Gothic or Old Saxon, and a mere hint of it in Old Norse, in the form of *skera tungu ór höfði manni* “cut the tongue out of man’s head” in Cleasby and Vigfusson’s (1957) Old Icelandic dictionary, where the dative marking in *manni* suggests the presence of the construction. However, its uniform presence in Germanic indicates that it is an original Germanic trait (although König and Haspelmath [1997] and Haspelmath [1999] treat it as an areal feature, such that contact may have played a role in its distribution within Germanic as well).

German:

- (4) Die Mutter wäscht dem Kind die Haare.  
 the mother wash.3S the-DAT child the-PL hair.PL  
 ‘The mother washes the child’s hair.’

Dutch:

- (5) Men heeft hem zijn arm gebroken.  
 IMP have.3S 3S-OBJ his arm break-PART  
 ‘They broke his arm.’ (König and Haspelmath 1997, 554)<sup>2</sup>

Frisian:

- (6) Ik stompte my de holle.  
I bump.PAST me.DAT the head  
'I bumped my head.' (Jarich Hoekstra, July 2001, personal communication)

Scandinavian:

- (7) Någon bröt armen på honom.  
someone break.PAST arm.DEF PREP 3S.OBJ  
'Someone broke his arm.' (König and Haspelmath 1997, 559)

Icelandic:

- (8) Han nuddaði á henni fætur-na.  
he massage.PAST on her.DAT leg.PL-DEF:ACC  
'He massaged her legs.' (ibid.)

Faroese:

- (9) Eg hoyrði røddina á honum.  
I hear.PAST voice on 3S.DAT  
'I heard his voice.' (Lockwood 1955, 105)

Yiddish:

- (10) (a) Di mame hot em gevasht di hor.  
the mother have.3S 3S.DAT wash.PART the hair  
'The mother washed his hair.'
- (b) Di mame hot gevasht di hor farn kind.  
the mother have.3S wash.PART the hair for.the child  
'The mother washed the child's hair.' (Jim Matisoff, July 2001, personal communication)

Old English had external possessor marking, as in:

- (11) þa cnitton hi rapas . . . hire to handum and fotum  
then tie.PAST they rope-PL her.DAT to hand.DAT.PL and foot.DAT.PL  
'then they tied ropes . . . to her hands and feet' (B. Mitchell 1985, 125)

However, the construction, already optional in Old English (B. Mitchell 1985, 126), decreases in frequency throughout the Old English period (ibid., 126–27; Ahlgren 1946). Visser (1963, 633) notes it as “common” but not obligatory in Middle English, but almost completely obsolete by the Modern English period. Modern English retains but sparse remnants of the earlier construction, as in *She looked him in the eyes* (König and Haspelmath 1997, 554, 560).

Ahlgren (1946, 201–2) suggests that English may have lost the external possessor construction due to the collapse of the dative and the accusative in English case marking. This early explanation has its echo in the emphasis in modern treatments on tracing historical developments in English to loss of overt case distinctions. But this surely cannot serve as an explanation for the loss of external possessor marking when Dutch and Scandinavian have experienced the same collapse of dative and accusative in pronouns and yet retain the feature. It is also germane that even a

language that does retain the dative/accusative contrast robustly, Icelandic, has nevertheless shed dative-marked external possessives for marking them with the locative. Obviously collapse of case marking was not a causal factor in English (cf. Haspelmath 1999, 125). Nor can Ahlgren's suggestion (1946, 210–16) that Latin was a deciding factor stand, when French and other Romance languages retain dative external possessor marking (*il m'a frappé la main* "he hit my hand") despite Latin playing as influential a role as a language of scholarship in their lifespans as it did in that of English.

Thus the question that arises is why it is ungrammatical in English to say "They broke him his arm," when Dutch has *Men heeft hem zijn arm gebroken*, or why English does not have "Someone broke the arm on him" as the Scandinavian languages do. Proposing a pathway of semantic evolution of prepositions upon which external possessives fall at a highly evolved point, Haspelmath (1999, 130–31) suggests that English does not use the preposition *to* in the function because it has yet to abstractualize to this extent. This is well taken in explaining why *They broke the leg to him* is ungrammatical, but does not account for why English does not instead encode external possession with oblique pronouns *without* a preposition, as Dutch does, or with locative prepositions as in Scandinavian. Under any internally based account, it is indeed "difficult to find a proper explanation," as Visser (1963, 633) puts it, of the absence of external possessor constructions in English; like so many, Visser is left to simply describe the change.

König and Haspelmath (1997, 583) note that the general tendency throughout Europe is for dative external possessives to recede, with only Baltic, Slavic, and Albanian preserving them as robustly as in early documents. Yet our question regarding English must be why it has lost the feature so quickly and thoroughly. Today, in Europe Welsh and Breton (and Turkish if we count it as a European language) are the only other languages that lack the construction, with even Finnish and Hungarian having picked up reflections of it, presumably through Sprachbund effects (ibid., 587–88).

Indeed, the loss of this construction in English must be viewed against concurrent cross-linguistic tendencies that are the very source of constructions such as these. Despite the broad trend over millennia to dilute external possessives, for instance, Romanian has extended the semantic boundaries of the "personal sphere" somewhat:

- (12) Ne-am            lăsat            bagajele    în autocar.  
 PL.DAT-have.PL leave-PART baggage.PL in coach.  
 'We left our baggage in the coach.' (Deletant and Alexandrescu 1992, 108)

Also germane here is what Icelandic linguists term "dative sickness," whereby over time, the marking of experiencers as dative rather than accusative is increasing rather than decreasing, as in:

- (13) (a) **Mig**        brestur    kjark.  
           1S.ACC lack.3S courage.ACC  
       (b) **Mér**        brestur    kjarkur.  
           1S.DAT lack.3S courage.NOM  
           'I lack courage.' (Smith 1992, 291)

Smith identifies dative sickness as symptomatic of a general diachronic tendency for case marking to decrease in what he terms “abstractness”; under his definition, abstractness decreases as the linking of grammatical relations (general) to semantic roles (specific) becomes more explicit.

Thus the pan-European tendency to dilute dative possessor marking coexists with a countervailing possibility that a grammar may also drift into increasing the overt marking of particular semantic roles applicable to a given general grammatical relation. English shunned the pathway that Romanian and Icelandic have taken, so decisively that external possessor marking vanished completely.

Finally, it is relevant that the only other Germanic language lacking external possessive constructions is Afrikaans, whose structure is now agreed upon to have been decisively impacted by extensive acquisition of Dutch as a second language by people of various ethnicities. This is one suggestion for the reason that the situation in English is not simply a matter of internally driven “business as usual.”

English, then, is unique among its European sisters in having chosen not to mark an inalienable, “personal” shade of experiencerhood. While just beyond the island both northward and eastward other Germanic languages have developed and retained a degree of inalienable possessive marking, English makes no grammaticalized differentiation between *He grabbed my hand* and *He grabbed my folder*, and as in the inherent reflexive marking case, is quite unusual in this even among European languages as a whole. Again, English is less overspecified than we would expect in this regard.

### 3.3. *Grammatical Gender Marking on the Article*

The mechanics of the loss of grammatical gender in English have been well covered, recent examples including Thomason and Kaufman (1988) and Lass (1992, 103–16). By the end of the twelfth century, grammatical gender was already all but lost in northern dialects; two centuries later, it had all but disappeared even in the south (cf. Strang 1970, 265).

However, in the emphasis on the phonological predictability of these changes, it has been less acknowledged that this change left English *the only European Germanic language with no grammatical gender marking*—despite the obvious vigor of phonetic erosion and analogy in its sisters. Moreover, this inflectional erosion also eliminated Old English’s declensional class marking.

Certainly, as Thomason and Kaufman emphasize, English is hardly unique in having lost grammatical gender marking on nouns themselves—Mainland Scandinavian, Dutch, and Frisian have only remnants of segmental indication of gender on nouns. Even Old English had already moved considerably in this direction, the emergence of multiple homophonies via phonetic change having already rendered the nominal morphology “relatively inexpressive and ambiguous,” as Lass (1992, 104) puts it and so many others have noted. The Germanic syllable-initial stress system is well known for having encouraged the erosion of unaccented word-final segments, which left nominal morphology especially vulnerable.

But nominal inflections were only a subset of the grammatical gender marking apparatus. To quote Lass again, in Old English “the richest and most distinctive

marking for nominal categories is on determiners, in the strong adjective declension, and in pronouns" (ibid., 106). Here, the determiners are particularly important; specifically, the articles. Concurrently with the erosion of the nominal inflections, an initial three-gender distinction in the demonstrative/article—*se* (masc.), *seo* (fem.) and *þæt* (neut.)—collapses into the gender neutral *the*.

Scholars on the history of English typically subsume the inflectional erosion and the collapse of the article's gender distinctions under a general "shift" from a grammatical to a natural gender-marking system. Such scholars include Strang (1970, 265, 268) and Lass, who describes this as a "cumulative weighting of 'decisions' in favour of natural gender." However, in a broader view, the "drift" characterization can serve only as a description rather than as an explanation. This is because it begs the question as to why all of the other Germanic languages of Europe, despite the erosion of the nominal inflections, maintained a grammatical gender distinction in the articles.

Surely, Dutch, Frisian, Danish, and Swedish collapsed the original Germanic masculine and feminine into a common gender contrasting with a neuter. But they only went this far, and it is unclear why English could not have done the same. It would even seem to have been phonetically plausible for *se* and *seo* to collapse into, perhaps, *se*, with *þæt* remaining as a neuter marker. Strang (1970, 268) states that "gender, as a grammatical system, can hardly survive the transformation of the personal pronoun system"—but since Dutch, Frisian, and Mainland Scandinavian underwent similar collapses in their pronouns, it is unclear that the disappearance of grammatical gender in English was so foreordained. (Note also that some non-standard Mainland Scandinavian dialects retain all three genders [e.g., Haberland 1994, 324].) Similarly, Lass's observation on "cumulative weightings" is obviously correct in itself, but reveals English speakers to have been unique among Europe's Germanic languages in this regard.

Is it possible that the phonetic shapes of Old Norse's articles led to the collapse of gender distinctions in Old English's cognate items as English speakers intermingled with Scandinavian settlers? It would seem that this explanation will only serve us but so well. The definite articles English speakers would have heard Old Norse speakers using were *þæn* (masc.), *þe* (fem.), and *þæt* (neut.) (Gordon 1927, 302). It is logical that English speakers in contact with Old Norse might have replaced the initial consonant of *se* and *seo* with [θ], the varieties being so typologically close that small adjustments like these could have gone a long way in easing communication. Yet contact *qua se* gives us no reason to assume that the immediate result would have been a single gender-neutral item *the*. When bilingualism between Western Danish and Low German was common in the thirteenth to fifteenth centuries, Danish lost much inflection (Haugen 1981) but retained a two-way gender distinction in its definite article (rather than reducing all forms to, for example, [də] on the phonetic model of what was common to the German definite articles). On the contrary, one plausible scenario is that English would have developed, for instance, a common gender item such as [θe] and preserved its neuter *þæt*. If such had occurred, there are no grounds for assuming that phonetic erosion would inevitably have eliminated the distinction in final consonant between the two forms. Faroese is a living demonstration, with its cognate configuration, masculine and feminine *tann* and neuter *tað*.

After centuries of regular use, the two remain distinct, partly due to the inherent conservativity of heavily used items. Only in English was this tendency overridden in favor of eliminating grammatical gender entirely.

Two broader observations highlight that English's lack of grammatical gender is a more "interesting" fact than generally assumed. First, among Germanic languages, again English's only parallel is Afrikaans, a language whose history was heavy with second-language acquisition.

Second, it has been seldom remarked that in its lack of any kind of grammatical gender within the noun phrase, English is unique not only among European Germanic languages but also among *all the languages of Europe*. As we would predict from the tendency for erosion and analogy to erase word-final morphology, there are scattered instances recorded in Europe of the loss of grammatical gender—but only in a few nonstandard dialects of particular languages, not all of the languages' dialects. Examples include Western Danish (Haberland 1994), Ostrobothnian Swedish, Tamian Latvian (Mathews 1956), and Mandres Albanian (Hamp 1965). Crucially, in the last two cases, gender was lost not through internal change but because of language contact (with Livonian and Turkish respectively)—significant for our argument that English was crucially affected by contact. To be sure, the two cases of internal loss are both Germanic. Yet the question remains: why is English the only European language in which *all* dialects have lost grammatical gender, such that today English speakers are the only Europeans who encounter grammatical gender marking as a new concept when acquiring another European language? Or, to view this from another angle, English is the only language in Europe where loss of grammatical gender occurred so quickly and completely as to be a fait accompli across all dialects by the time European languages were being standardized in the middle of the last millennium.

Thus English contrasts with its sisters in lacking a particular type of overspecification and structural elaboration: the obligatory marking of noun phrases according to categories generally only marginally correspondent with any real-world distinctions, complete with its attendant irregularities. Some linguists are uncomfortable with the idea of grammatical gender marking as "useless," and indeed once present, it can be useful in reference tracking (e.g., Heath 1975; Foley and Van Valin 1984, 326). However, Trudgill (1999) rather conclusively refutes the notion that the gender marking arose *in order to* serve such functions, it being rather clear that the markers can be "exapted" into this function along the nonteleological lines of Keller's (1994) "invisible hand" concept. As such I am inclined to classify grammatical gender as equivalent to, in Lass's deft phrasing, "linguistic male nipples" (1987, 13).

Yet even if one is inclined to disagree, the point stands. Whether English was the only Germanic language disinclined to preserve a useless feature, or the only one that shed the feature despite its being useful to communication, it has been unique in its "streamlining" orientation.

### 3.4. Derivational Morphology

As is well known, English is unusual among Germanic languages in the volume of original Germanic derivational morphology that it has lost: the rich Proto-Germanic

battery of affixes (cf. Voyles 1992, 270–79) has been reduced to scattered remnants. The typical account focuses on the frequent replacement of Germanic derivational affixes with French ones. But some analysts note that the loss appears to have predated significant contact with French (Strang 1970, 191; Dalton-Puffer 1995, 39); while French lexical items often only appear in texts after the Norman occupation, Hiltunen (1983, 92) describes the derivational loss as virtually complete as soon as Middle English texts begin. This means that in the strict sense, Old English apparently simply let a great deal of its derivational apparatus go; the French replacements were a later consequence of geopolitical developments.

### 3.4.1. VERB PREFIXES

Of course, in the case of many of the Old English prefixed verbs, we cannot speak properly of loss given that they were simply replaced by equivalent phrasal verbs, as Hiltunen describes: *toberstan* > *to break apart*, *inlædan* > *to bring in*, and so on. However, this only happened where the prefix either coexisted with a free preposition (e.g., *in*) or was of semantics robust and discrete enough to be readily substituted by an equivalent free word: *down* for *nifer-*, *around* for *ymb-*, *up* and *out* for the intensificational uses of *for-*, as in *forbærnan* > *to burn up*.

This process would then fall under the rubric of the general loss of morphology in the inflectional realm, where similarly, distinctions encoded by affixation are often replaced by ones encoded by free morphemes (e.g., prepositions) or word order. But all of the derivational losses cannot be subsumed under a simple substitution of the analytical for the synthetic, as certain prefixes were instead eliminated from the grammar without any substitution. Namely, the prefixes with semantics leaning more toward the grammatical, abstract pole simply disappeared except for fossilized remnants. Thus English lost its transitivity *be-* (*seon* “to see,” *beseon* “to look at”), and *ge-*, alternatively described as transitivity (*Visser* 1963, 127) or perfectivizing (*Mitchell and Robinson* 1986, 58), as in *ærnan* “to run,” *geærnan* “to reach, attain by running” *winnan* “to toil,” and *gewinnan* “to conquer.”

Another question arises with another use of *for-*. Grammars typically describe its relatively compositional uses, such as the intensificational one. However, a survey of its uses across the Old English lexicon shows that its contribution had often bleached to the point that there was little or no perceptible difference of meaning between the bare verb and its conjunction with *for-*, as in *helan/forhelan* “to conceal” and *polian/forpolian* “to lack, be deprived of.” This is also the case in German, Dutch, Afrikaans, Yiddish, Mainland Scandinavian, and Frisian, where there are doublets of this type whose differentiation of usage is, at best, highly subtle and sometimes register-bound (German *sterben/versterben* “to die”). It is likely that this was a step toward the reanalysis of the *for-* cognate in these languages as simply a marker of verthood, connoting transformation, and in this usage extended to nouns and adjectives as well as verbs: German *verlängern* “to make longer,” Dutch *vernederlandsen* “to Dutchify,” Afrikaans *verafrikaans* “to Afrikaansify,” Swedish *förgifta* “to poison,” and so on. Even heavy contact has not hindered this development, as we see in Afrikaans as well as Pennsylvania German, both threatened by English, where nevertheless the cognate prefix is used to create new verbs: *fərbötft* “all botched up” (Van



Ness 1994, 433). Old English could, theoretically, have replaced the compositional uses of *for-* (such as the intensificational one) with phrasal verb particles, but otherwise retained *for-* as a verbalizer of this kind as many of its sisters did. But instead, while the intensificational uses were indeed replaced by phrasal verbs, in its other uses *for-* simply disappeared as a productive morpheme.

In short, English's loss of prefixes entailed a significant degree of loss of overspecification of constituent class and semantic nuance. As it happens, Icelandic and Faroese have also opted for phrasal verbs to the virtual exclusion of the Germanic verbal prefixes, including losing *be-*, *ge-*, and *for-* cognates. More precisely, it was Old Norse that shed these prefixes starting in the mid-twelfth century (as well as some derivational suffixes), which also meant that Mainland Scandinavian also lacked them at its emergence, and only borrowed them later from Low German, along with a battery of lexical items comparable in volume to the French component in English (cf. Braunmüller 2002, 1037–38). This means that my marking the Mainland Scandinavian languages as having the derivational prefixes in the table is, strictly, only a synchronic fact.

However, my thesis hardly rules out that a given Germanic language other than English—such as, in this case, Old Norse—might have also shed a given feature, due either to developments elsewhere in the grammar or to sheer chance. However, my thesis is indeed that to couch the developments in English as unremarkable results of a “trend toward analyticity” misses a larger point. In that light, we must first note how very much *inflectional* morphology Icelandic and Faroese nevertheless retain, as well as the various other overspecified features discussed thus far. The Mainland Scandinavian languages have retained much less inflectional morphology—so much less, in fact, that if English did not exist, an analyst might focus on them (or Dutch and Frisian) as strangely low on inflection amid Germanic (it is not an accident that Kusters [2003] devotes a chapter to Mainland Scandinavian languages as markedly simplified inflectionally). Yet Swedish, Norwegian, Danish, Dutch, and Frisian all retain the features discussed above, as well as others to be discussed below—a stark qualitative contrast with English remains clear.

Moreover, it is also relevant that even Afrikaans, despite its heavy inflectional loss, has retained the verbal prefixes *be-* (*beslis* “to decide”), *ont-* (*ontken* “to deny”), and *ver-* (*verpletter* “to smash”), and not in fossilized form, but with a certain degree of productivity, as demonstrated by neologisms such as *beplan* “to plan,” *ontlont* “to defuse,” and *verafrikaans* “to Afrikaansify” (Ponelis 1993, 556–57).

That is, the derivational prefixes were lost in a grammar otherwise heavy with morphology, and preserved in a language that otherwise underwent major reduction in morphology. The loss in English appears traceable to something other than random phonetic erosion; something more specific was at work.

### 3.4.2. SUFFIXES

The losses of various derivational *suffixes* in Old English is overall less indicative of any trend against overspecification. Old English documents capture the language at a stage where some suffixes are robustly productive while others are falling by the wayside, some obscured from perception by the transformations of the root they



caused (most notably *-the*, as in *foul/filthe*, *young/you(ng)the*), others losing out in Darwinian competition with ones filling similar spaces (*-reden* and *-lac* versus *-ness*). In any language we see affixes at various points on the cline between glory and oblivion, and thus it is not necessarily the case that these losses in themselves suggest that Middle English was a “recessive” language as Dalton-Puffer (1995) proposes.

However, there is one case here which again points to a trend toward underspecification. English loses its infinitive marker *-(e)n* by the end of the 1400s (Lass 1992, 98). Other than English, only Afrikaans, Germanic’s contact language par excellence, has lost all signs of this marker completely. But in itself, the loss in English entailed no loss of overspecification. Already in Old English, there was a semantic difference between bare infinitives and those occurring with *to*, the former’s conjunction with the preceding verb connoting more transitive actions processed as one event (Callaway 1913 cited in Fischer 1997). When the infinitive marker wore off and all infinitives were then marked with *to*, this distinction was not lost: *-ing*-marked verbs came in to connote the more transitive relationship with infinitives now processed as less so (e.g., *I saw him doing it* versus *I wanted to do it*) (Mittwoch 1990; Fischer 1997, 126).

For the purposes of this argument, however, it is important to note that English lost not only the verb-marking reflex of this morpheme, but also those that served derivationally as word-building devices. In Old English, *-sian*, *-ettan*, and *-læcan* served to make nouns and adjectives into verbs (*ricsian* “to reign,” *licettan* “pretend,” *geanlæcan* “to unite” [Mitchell and Robinson 1986, 60]). These were flushed away: English did not take the route of, for example, letting *-(e)n* erode but retaining markers *-s*, *-ett*, and *-læc* as derivational equipment (e.g., retaining enough pairs like *clean/cleanse* for the suffix to be processable and even productive). To be sure, cognates of these suffixes are now defunct in English’s sisters. However, they all retained machinery elsewhere that overtly marks analogous derivational distinctions. Meanwhile, English instead chose the pathway toward its now notorious fondness for zero-denominal and zero-deadjectival derivation. In German, *I telephoniere*; in English I just *telephone*. The German cannot “get away with” “*ich telephon*,” nor is the cognate equivalent possible for speakers of any Germanic language but English.

### 3.4.3. DESCRIPTION VERSUS EXPLANATION

The causes traditionally adduced for this rather striking sloughing away of derivational apparatus within a few centuries leave more questions than answers. The idea that lack of stress rendered the morphemes uniquely vulnerable addresses a tendency rather than an inevitable death sentence. German and others, after all, retain, for example, past participle marker *ge-*, and even southwestern English dialects retained its cognates *y-* and *a-* (cf. W. Barnes 1886, 27–28 on Dorset) as participial markers until pressure from the standard rendered them extinct. This last questions Marchand’s (1969, 130–31) suggestion that the vowel-initial prefixes in particular were uniquely vulnerable, as well as the idea that erosion is the sole reason that the almost three dozen Old English verbs transitivized by *ge-* that Visser (1963, 127) lists were shorn of their valence markers.

Authors also sometimes suppose that an affix was ripe for elimination because it had many meanings (Dalton-Puffer 1996, 179, on adjectival marker *-ly*), or because its contributions to many stems were no longer semantically predictable (Marchand 1969, 130–31, on *for-*). However, in any language, a given affix may remain in productive use in a core meaning while its contributions to myriad roots have drifted into noncompositionality. The noncompositional uses are not evidence of imminent demise of the affix, but merely indications that the affix has been in use for a long time. The German *ver-* is a useful example. One usage conveys the notion of “away”: *jagen* “to hunt,” *verjagen* “to chase away.” There are extended meanings from this one, such as error (“away” from the right path), creating antonyms such as *lernen* “to learn” versus *verlernen* “to forget.” Meanwhile, many uses of *ver-* are unattributable to any of these meanings and must be learned by rote, such as *nehmen* “to take,” *vernehmen* “to perceive.” Yet this is not taken to signal that *ver-* is on its way out of the grammar; on the contrary, it is used productively to create new verbs (*verschlagnen* “to file under a subject heading” [Ingo Plag, personal communication]), with the non- and semicompositional results of its historical legacy simply dragged along by speakers. Thus the question is why English does not drag along noncompositional cases like *forbid* and *forgive* at the same time as creating words like “forenglish” to mean “to Englishify.”

Finally, there are explanations such as Visser’s (1963, 134) that a given affix disappeared because a great number of the words displaying it “dropped into disuse.” The implication would seem to be that the massive incursion of French words eliminated so much of the original Germanic lexicon that in some cases too few uses of a given affix remained to be processable by speakers. Yet it is well known that French words often took their place alongside Germanic words to create synonyms, often occupying different registers.

Visser’s list of now lost words where *be-* was affixed to roots still used in bare form in Modern English is worth citing in full: *bebark*, *bedwell*, *bechirp*, *beflow*, *befly*, *begaze*, *beglide*, *beglitter*, *bego*, *behoot*, *beleap*, *belie*, *bemew*, *berain*, *beride*, *berow*, *beshite*, *beshriek*, *besit*, *bescramble*, *bescratch*, *besparkle*, *beswink*. First, why could a healthy subset of words of this kind not have persisted alongside French equivalents, as *help* persisted alongside *aid*, and so forth? Certainly we would expect some to vanish by the sheer dictates of serendipity—but so many that today the prefix occurs on too few words to be processable to any but highly literate modern speakers? Even if all of the words on Visser’s list did for some reason “drop into disuse” by chance, why did speakers not come to apply the native affix to borrowed words, as they went on to apply borrowed affixes to native words (*speakable*, *bondage* [Dalton-Puffer 1996, 221], or today, *faxable*)? An alternate interpretation of the disappearance of the words is that it was the *affix* that speakers were rejecting, not the words themselves—especially when they so often retained the root itself (i.e., in reference to Visser’s list, *bark*, *dwell*, *chirp*, *flow*, etc.).

Thus the loss of derivational morphology in English takes its place alongside the loss of inherent reflexive marking, external dative possessor marking, and grammatical gender marking in rendering the language less accreted with complexity. In this case, English shed overt marking of transitivity with the jettisoning of *be-* and one use of *ge-*, and overt marking of noun- and adjectivehood with its shedding of derivational uses of reflexes of the infinitive marker *-(e)n* and the prefix *for-*.

3.5. *Directional Adverbs*

Germanic languages typically distinguish forms of adverbs of place according to location, motion toward, and motion away from (Swedish *här* “here,” *hit* “to here,” and *härifrån* “from here” [Holmes and Hinchliffe 1997, 115–16]). This was a Proto-Germanic feature (cf. Voyles 1992, 242).

Old English originally toed the Germanic line here (e.g., *her*, *hider*, *heonan*), but the system was already fragile, with *her* often used for motion toward, *-an*-suffixed forms losing their sense of “from” and being used as mere locationals, and so on (B. Mitchell 1985, 476; and Meroney 1945, 386, cited in B. Mitchell). This uncertainty and variability can even be taken as a suggestion that in the spoken language the distinctions were even more fitfully observed. In any case, even in its written form the system was already what Meroney (1945, 386) describes as “a stage of compromise between Germanic and Modern English,” and by the latter stage, English had become the sole Germanic language not to attend regularly to this distinction. Given that these forms were widely used in high literary English into the 1800s, it is difficult to place exactly when they passed out of spoken English. However, all would agree that they are no longer current today beyond frozen expressions like *hither and yon* (which, to this writer’s ear, are restricted for modern generations to the ironic, at least in the United States).<sup>3</sup>

The loss of the “motion away from” forms did not in itself lead to a loss in encoded meaning, since the word *from* was recruited to serve the same purpose: *heonan* became *from here*, and so on. However, motion *toward* a destination is often contained within the semantics of a verb of motion, and in these cases, English, as so often elsewhere, took the route of leaving the nuance to context.<sup>4</sup>

Table 4.2. Directional Adverbs in Germanic

Eng	OE	Ger	Du	Fr	Yi	Sc <sup>5</sup>	Ic	Fa	Afr
here	her	hier	hier	hjr	hi	här	hér	her	hier
	hider	her(-)	hiernaartoe	hjrhinne	aher	hit	hingap	higar	heimatoe
there	þær	dort	daar	dêr	dort	där	þar	har	daar
	þider	hin(-)	daarnaartoe	dêrhinne	ahin	dit	þangap	hagar	daarnatoe
where	hwær	wo	waar	wêr	vu	var	hvar	hvar	waar
	hwider	wohin	waar...heen	wêrhinne	vuhin	vart	hvert		waarheen

Certainly as directionality itself goes, the overt distinction is not entirely foreign to English, as in *in the house* versus *into the house*. Moreover, even with the adverbs themselves, colloquial Englishes often make the distinction variably with *where* as in *Where is she at?* versus A: *We’re going now*. B: *Where to?* But this is hardly the case with most applicable adverbs, and the point remains that it is obligatory in the standard *not* to mark the distinction on any adverb: *hither*, *thither*, and *whither* are strictly archaic words foreign to even the highest registers of Modern English. Moreover,

arguably, the *absence* of the distinction is grammatical to all English speakers—there is probably no Anglophone context on earth where asking *Where?* rather than *Where to?* would sound non-native or clumsy. *Thence* is used occasionally in nonfiction prose, but as a highly formal word, and in its pragmaticized temporal meaning rather than its original directional usage.

Other Germanic languages differ slightly in the degree to which the distinctions are obligatory. The Mainland Scandinavian varieties differ among themselves on this, for example. Tiersma's Frisian grammar has the use of *hjr* with the *come*-verb as grammatical; Faroese even has an outright gap with *hvar*. The distinction also varies in terms of scope of application within the grammar (German's conventionalization of the conjunction of *her-* and *hin-* with prepositions and verb particles being an extreme), and in where particular usages fall in terms of register. But in all of the languages, the distinction is a robust aspect of their grammars (to my knowledge, even nonstandard varieties), usually applying to a wide range of adverbs.

We cannot simply classify this loss as a mere symptom of the erosion of morphology in Old English. For one, the "motion" forms in Old English differed in shape from the locational reflexes far beyond the affix itself. If morphological loss were the smoking gun here, then we might expect, for example, *hid* and *heon* to have resulted, still distinct from *her* (> *here*). In any case, too often Germanic languages have maintained this distinction *despite* vast morphological losses: Afrikaans is the most pointed demonstration, followed by Mainland Scandinavian and Dutch.

But English alone shed these forms, and Sapir (1921, 169–70) artfully parsed the grammarwide developmental impetus that this demonstrated:

As soon as the derivation runs danger of being felt as a mere nuancing of, a finicky play on, the primary concept it tends to be absorbed...[an] instance of the sacrifice of highly useful forms to this impatience of nuancing is the group *whence, whither, hence, hither, thence, thither*. They could not persist in live usage because they impinged too solidly upon the circles of meaning represented by the words *where, here* and *there*. That we add to *where* an important nuance of direction irritates rather than satisfies.

Crucially, in Germanic only English speakers felt such an "irritation." Alone, this anomaly in English could be seen as a mere fluke; in conjunction with the previous four cases adduced, a larger process reveals itself at work.

### 3.6. *Be with Past Participles*

A hallmark of Germanic (and Romance) is the use of the verb *to be* with a large subset of intransitive verbs in the perfect: German *er hat gegessen* "he has eaten," "he ate"; *er ist gekommen* "he has arrived," "he arrived." Of course, the precise domain of intransitives to which *be* applies varies across the languages, but the basic distinction is retained even in Afrikaans (*Ze zijn vertrokken* "they have left" [Ponelis 1993, 444]) and Yiddish (*Ikh bin geblibn* "I stayed" [Lockwood 1995, 83]). The virtually uniform distribution of this feature suggests Proto-Germanic inheritance; it is found as early as Old Norse (Brenner 1882, 129) and Old Saxon (Ramat 1998, 403). Accordingly, Old English marked this distinction with the verbs *beon* and *wesan*:

- (14) hu    sio    lar            Lædengeðioðes            ær            ðissum    afeallen            wæs  
 how the learning Latin-language.GEN before this.DAT fall-away.PART was  
 ‘How the learning of Latin was fallen away before this.’  
 (Mitchell and Robinson 1986, 111)

Yet as so often with typical Germanic constructions, already in Old English the usage was apparently in flux, with *habban* encroaching on the domain of *beon* and *wesan*. Bruce Mitchell (1985, 302–4) suggests that none of the attempts over the years to delineate a principled semantic distinction between the use of a verb with *habban* as opposed to *beon* or *wesan* withstand scrutiny, and questions whether the documentation even indicates a grammaticalized *be*-perfect, as opposed to a typical use of a *be*-verb with stative adjectivals.

By the 1500s, the use of *be* in the perfect had largely shrunk to the change-of-state class of intransitives such as *come*, *become*, *arrive*, *enter*, *run*, and *grow*:

- (15) *And didst thou not, when she was gone downstairs, desire me to be no more so familiarity with such poor people?* (Henry IV, II.i.96) (cited in Traugott 1972)

(Compare also Rissanen [1999, 213]). In their variationist analysis based on texts as representative as possible of the spoken language, Rydén and Brorström (1987, 200) show that by the early 1800s, overall usage of *have* over *be* surpassed the 50 percent mark, with prescriptive grammarians granting tolerance of the *be*-perfect to an ever narrower class of verbs over the century (206–11). Today the usage has vanished except for those in frozen form with *go* (i.e., if I may, *The construction is gone*).<sup>6</sup>

The pathway English followed is striking given that the development of *be*-perfects was an innovation in Germanic and Romance rather than an inheritance from Proto-Indo-European, and in many languages the domain of *be* has spread rather than contracted over time—compare the varying extents of its application across Western European languages in Sorace (2000). English instead reversed this pathway of overspecification of intransitivity, thus joining the elimination of inherent reflexive marking and the derivational prefixes *be*- and *ge*- in rendering English the Germanic language with the least overt marking of valence.

Obviously we cannot lay this change at the feet of inflectional loss, nor can it be subsumed under the rubric of the drift toward analyticity.

Typically, the disappearance of the *be*-perfect is attributed to the recruitment of *be* as a marker of the passive (Mustanoja 1960, 501; Traugott 1972, 145; Mitchell 1985, 299; Rissanen 1999, 213). There is even comparative support for this explanation, in the fact that Swedish, the only Germanic language other than English that lacks the *be*-perfect (*Vi har rest till Spanien förr* ‘We have gone to Spain before’ [Holmes and Hinchliffe 1997, 100]), has also recruited its *be*-verb *vara* to mark the passive (cf. Rissanen 1999, 215).

But the causal relationship here is not absolute. Icelandic, too, forms its passives with its *be*-verb *vera* (*ég var barinn* ‘I was hit’) and in the perfect uses *have* with both transitives and intransitives (*ég hef komið* ‘I have come’) (Kress 1982, 148–49). Yet Icelandic also uses *vera* with intransitive verbs of motion and change-of-state to con-

note the resultative: *ég er kominn* “I am come, I am here,” and the class of verbs used this way is large (ibid., 152–53)—the strategy is by no means the recessive, marginal archaism that it was, for example, by Early Modern English.

It also bears mentioning that Swedish is unique even in Mainland Scandinavian in lacking a *be*-perfect. Danish and Norwegian retain it: Danish *Barnet er kommen* “the child has come” (Thomas 1991, 133–34), Norwegian *Han er reist* “he has left” (Strandskogen and Strandskogen 1986, 21). This highlights the general tenacity of this feature in a language once it arises (although it happens to be receding in Norwegian). Yet note that English *parallels* the Swedish exception—as always, if a Germanic language other than English happens to opt for context where family tradition calls for being explicit, English will have done the same (the Faroese lack of a *to*-marked *where* being another example).

### 3.7. Passive Marking with Become

Another Germanic tribal marker is the use of a verb “become” to form the passive (German *Die Tür wird geschlossen* “The door was closed”), this including Afrikaans with its *word* (*Die trui word gebêre* “The jersey is put away”). Uniform distribution again suggests a Proto-Germanic pedigree, with its presence as far back as Old Norse (Heusler 1950, 137) and Gothic (Streitberg 1906, 182–83) reinforcing the reconstruction. While Swedish indeed uses its *be*-verb *vara* in the passive, it does so in conjunction with its verb “to become” *bli*; *vara* conveys a “stative” passiveness in line with its semantics: *Himlen är täckt av moln* “The sky is covered in cloud,” while *bli* conveys more perfective semantics: *Han blev påkörd av en bil* “He was run down by a car” (Holmes and Hinchliffe 1997, 109). Icelandic, too, retains its *verða* along with its use of *vera* “to be” in the passive, in a division of labor in which *verða* is the marked, but hardly marginal, member (Kress 1982, 150).

As frequently, Old English followed the Germanic pattern in already rather atrophied fashion. *Beon* and *wesan* were already easing out *weorþan* in the passive, and Mitchell (1985, 324–35) rather spiritedly refutes common claims that this was instead a regularized distinction between perfective semantics conveyed by *weorþan* and stative ones by *beon* and *wesan* (along the lines of Swedish’s *vara* and *bli*). Mitchell argues that all of the forms were used with both readings, but Denison (1993, 418–19) and Kilpiö (1989) show that more properly, *weorþan* was restricted entirely to the actional while *beon* and *wesan* were grammatical in both this and the stative meanings. By Middle English, the *weorþan*-passive is not just recessive, but nonexistent (Rissanen 1999, 325).

Modern English has innovated the marking of passive with *get* (*He got hit*) and *have* (*He had his hair cut*), but both are pragmatically constrained, encoding especial activeness on the part of the subject, with the *have*-passive essentially a causative. Overall, English remains the only Germanic language without a lexical item dedicated to expressing a pragmatically neutral manifestation of the passive. Only in Icelandic is it even grammatical to use the *be*-verb to say *He was kicked*, and even it has retained *verða* alongside. Meanwhile, properly speaking, Swedish has recruited *vara* into a *subdomain* of the passive, retaining *bli* to distinguish the “true” passive as

opposed to its more stative manifestations. Once again, English opts for underspecification where its sisters insist on dotting the *i*'s and crossing the *t*'s.

### 3.8. V2

All Germanic languages but English have verb-second word order, including Afrikaans. This is generally agreed to be a Proto-Germanic feature (Hopper 1975, 82; Ramat 1998, 410–13).

The languages differ in their particular manifestations of the phenomenon, often classified as “asymmetric” when V2 occurs only in root clauses and “symmetric” when V2 occurs in both root and subordinate clauses. Which type of V2 Old English manifested is disputed (Van Kemenade 1987 versus Pintzuk 1991), but it is uncontested that it was a V2 language:

- (16) On twam þingum hæafde God þæs mannes sawle gegodod.  
 in two.DAT thing.DAT have.PAST God this.GEN man.GEN soul endow.PART  
 ‘God had endowed this man’s soul with two things’ (Fischer et al. 2000, 107)

V2 in English begins a decline in the fifteenth century, and is essentially dead by the seventeenth (Jacobsson 1951; Nevalainen 1997). The question obviously arises as to why.

One current consensus links the loss to the erosion of verbal inflectional morphology. A general assumption is that V2 results from verb movement, specifically to C (Den Besten 1983), and inflection-based accounts of V2 loss suppose that the erosion of verbal morphology led to the verb staying in place rather than moving upward in its clause (e.g., Fischer et al. 2000, 135–36). But overall, the explanations offered in this case lack explanatory power or falsifiability.

For example, an inflection-based account of the loss of V2 presumes that the very small difference in degree of verbal inflection between Mainland Scandinavian and English determined that the former would preserve V2 while the latter would lose it. Yet this difference consists only of the fact that Mainland Scandinavian marks the present in all persons and numbers with *-r* (Swedish *jag arbetar* “I work”) while English inflects in the present only the third-person singular. This would appear to attribute a profound configurational transformation to a rather minor discrepancy, especially given that discourse studies show that the third-person singular is by far the most frequent in speech (e.g., Greenberg 1966a, 45), such that the inflected form in English constitutes a disproportional component of input to learners. Where is the cutoff point that determines how “weak” inflection must be before it conditions a change in movement rules?

This question is all the more pressing given that in reference to a related process, Roberts (1993), Rohrbacher (1999), and others have argued that loss of verbal inflection in both English and Mainland Scandinavian led to the loss of verb movement to I in subordinate clauses. A demonstration case is English, in which previously the verb moved ahead of the negator, adverbs, and other elements, as in . . . *if I gave not this accompt to you* from 1557 (Görlach 1991, 223). But there appear to be no principled accounts to date that motivate the differing fates of V-to-I and V-to-C



movement in English. Precisely why did inflectional loss preserve V2 in matrix clauses but eliminate V-to-I in subordinate clauses in Mainland Scandinavian, while eliminating *both* movement processes in English?

We might be tempted to suppose that for some reason, the small difference in degree of inflection was indeed responsible for the very specific effect of preserving V2 but not V-to-I movement in Mainland Scandinavian. But then the latest evidence suggests that the decisive causal link is solely between “strong” inflection and verb movement; when inflection is “weak,” then the verb may or may not move (I. Roberts 1999, 292). The Kronoby Swedish dialect preserves V-to-I despite the inflectional erosion (Platzack and Holmberg 1989). Meanwhile, Kroch and Taylor (1997) argue that when verbal inflection eroded in English dialects in the north (under Scandinavian influence), the result was not the loss of V2, but a mere change in its configuration, from symmetric (the authors assume Pintzuk’s analysis) to asymmetric. Baptista (2000) shows that there is evidence of verb movement in Cape Verdean creole despite its having but a single verbal inflection. Rohrbacher (1999) notes that Faroese verbs do not move to I despite robust plural inflection on verbs. He thereby surmises that “strong” inflection entails overt marking of the first and second persons in at least one number of at least one tense. But this stipulation is rather *ad hoc*, contradicting the centrality of the third-person singular in discourse, and would seem to have been invalidated by dialects like Kronoby Swedish. It would appear that, simply, the correlation between inflection and verb movement is a rather loose one. While work on the relationship of overt morphology to verb movement continues to be refined (e.g., Bobaljik and Thráinnsson 1997), it seems clear that the link is too weak in itself to offer a conclusive explanation for what happened in English in comparison to its sisters.

Lightfoot (1997, 268–69) argues that this kind of gap in explanatory power is not problematic for generative diachronic syntacticians. In his view, their enterprise is strictly to use language change to illuminate the effects of synchronic parameters, and under this constrained conception, the reasons for the changes are irrelevant: “Sometimes the concern with explanation is excessive . . . such things happen for various reasons which are often of no particular interest to grammarians.” Thus Lightfoot prefers to simply chart changes like the loss of V2 in terms of input gradually depriving learners of “triggers” motivating the setting of the appropriate parameters.

However, the reason for the disappearance of the “trigger” for V2 is crucial to this particular thesis, and I suggest that the reason is less obscure, or “contingent,” as Lightfoot (1997) has it, than it might seem. In becoming the only Germanic language without V2, English opted for what can be argued to be the less complex syntactic configuration. Despite its air of “linguocentricity” when argued by an Anglophone, there is a great deal of evidence that SVO is a universally unmarked order. Kayne (1994) is an articulate generative demonstration. Pidginization and creolization data also support SVO as a “universal” order. Creoles tend to be SVO regardless of the word order of their substrate languages, such as Berbice Dutch Creole, formed between speakers of Dutch and the SOV Niger-Congo language Ijo (Kouwenberg 1994).



Linking the disappearance of V2 to a decomplexifying imperative also sheds light on another analysis of English's unique treatment of this feature: Paul Kiparsky's (1995). Kiparsky proposes that the development of COMP in both subordinate and matrix clauses was an innovation in Proto-Germanic, Proto-Indo-European being presumably a more clausally paratactic grammar (cf. E. Hermann 1895). Kiparsky reconstructs that while all other Germanic languages conventionalized COMP in matrix clauses, English was unique in first having COMP as only optional in Old English, and then eliminating it. He notes that Old English is unique in Germanic in allowing matrix clauses such as the one below, where the verb does not raise, while equivalent sentences were unattested in other early Germanic languages like Old High German and Old Norse:

- (17) He þa his here on tu todælde.  
 he then his army in two divide.PAST  
 'He then divided his army in two.' (Orosius 116.16) (P. Kiparsky 1995, 143)

But Kiparsky's analysis begs the question, like so many accounts of the history of English, as to just *why* English was unique in this regard. Kiparsky is concerned with diachronic syntactic analysis within the Principles and Parameters framework, and thus presumably concurs with Lightfoot regarding the theoretical import of such questions. However, within a frame of reference where such questions are more urgent, we see that English, in failing to develop an obligatory COMP node in matrix clauses, would once again be opting for the less complexified path than its sisters, eschewing V2 in the vein of most of the world's languages while its sisters drifted into a typologically unusual quirk.

Importantly, Kiparsky notes that the presence of COMP in matrix clauses in Old English is variable rather than absent. But as with inherent reflexives, we must ask why English did not conventionalize rather than eliminate the feature. Old Norse did; by the time it is documented in about 1100, it has already grammaticized COMP in matrix clauses. Old High German was similar. Meanwhile, only in Old English does the variability persist—and then eventually yield to dissolution.

In sum, in eliminating V2, English eliminated a crucial structural elaboration that required the operation of a rule moving the verb to C, a feature whose supplementary character in general is illustrated by the typological rarity of the V2 feature beyond Germanic. Other Germanic languages held on to V2 despite inflectional erosion as rampant as that in English, namely, Mainland Scandinavian and Afrikaans. Inflection-centered, syntax-internal accounts have yielded stimulating explanations proper for assorted *variations* upon the manifestation of V2. But for the complete elimination of verb movement, the best they have provided to date are loose correlations. No amount of refinements of this framework has to date been able to explain why *only English shed V2 in affirmative sentences altogether*. I suggest, along with Danchev (1997), that only a larger, contact-based explanation can surpass this obstacle.

### 3.9. *The Disappearance of Thou*

By the 1700s, the originally plural *you* had replaced *thou* in standard English. Research on court documents from the northeast suggests that in spoken English, *you*

was already the conventional second-person singular form as early as the late 1500s, with *thou* used only in particular marked contexts (Hope 1994). Hope suggests that the wider use of *thou* in literary sources such as Shakespeare may have been a conservatism that spoken English had moved beyond.

This development is typically discussed within the larger context of the use of second-person plural pronouns in formal address to single persons across Europe. But Strang's (1970, 139) comment that "such a use, once introduced, must snowball," while obviously apt, does not explain why the "snowballing" went so far in English as to leave it the only Germanic language which lost a distinct second-person singular pronoun altogether. The usual result of the well-known development of "T-V" forms was for the V form to encroach ever more upon the realm of familiarity—but all of the other Germanic languages nevertheless retain the familiar form. If anything, the modern development has been toward the reassertion of the T form within the democratizing ideological tendencies of the post-Enlightenment age. Yet during just this period, English relegated *thou* to the archaism of the religious register. As Strang notes (140), it might not have persisted even here if the King James Bible had not happened to reproduce to such an extent the usage of Tyndale, who wrote in the early 1500s when *thou* was still in current use.

Clearly, neither inflectional loss nor a drift toward analyticity were related to a change which did not transpire even in Afrikaans, in which both of those processes were rife in its development. Furthermore, even highly isolating languages rarely display an isomorphy between singular and plural pronouns in the second person. Even the inquisitive undergraduate is often given to ask, when exposed to the T-V pronoun issue, why English went as far as to eliminate *thou* entirely while German retained its *du*, and so on. Often the professor can only offer an authoritative shrug.

But in fact, this development correlates with the eight we have seen so far in rendering English less overspecified than its sisters. Obviously the lack of a number distinction in the second person occasions no significant communicational difficulties. However, every single Germanic language but English has preserved this distinction, frill though it is, to the present day. To be sure, Dutch and Frisian have lost the *du*-cognate itself, but have nevertheless "exapted" other material to maintain a T-V distinction. English stands as unique among its sisters in having eschewed the nuance altogether.

### 3.10. *The Disappearance of Man*

English began with the usual Germanic endowment of an indefinite pronoun *man*, grammaticalized enough to have eroded phonetically to *me* by Middle English:

- (18) Ac **me** ne     auh to bien hersum     bute     of gode.  
       but one NEG ought to be     obedient except in good  
       'But one should not be obedient except in good things.' (Rissanen 1997, 520)

The pronoun appears to trace to Proto-Germanic. The only Germanic languages lacking a distinct indefinite pronoun are Icelandic and Faroese (Icelandic recruits *maður* "men" [Kress 1982, 113], while Faroese uses *man* but only as a Danicism, preferring to use *tú* "you" and *teir* "they" like English [Lockwood 1955, 125]). But *man*

is present (although variably) in their ancestor Old Norse (Heusler 1950, 147), suggesting that its eclipse in two daughters was a subsequent development. Meanwhile, Afrikaans does not retain Dutch's *men*, but instead uses the colloquial Dutch 'n *mens* (< "a person")—and the tendency is to shorten this to *mens*, creating what Ponelis (1993, 224) analyzes as a new pronoun.

As always, where a few of its sisters eliminated a feature English followed suit: English's *man* rapidly disappeared, essentially gone in the written language by the late fourteenth century. To the extent that it appears in regional speech after this, it is marginal, such as in the early 1900s in Cumberland as attested by Brilioth (1913, 111). The original form *man* had split off to connote "a human being":

- (19) panne **man** forget that he seien sholde  
 when one forget what he say. INF should  
  
 panne beð his tunge also hit cleued were.  
 then be.3S his tongue as it stuck were

'When a person forgets what he should say, his tongue is as if it werestuck.'  
 (Rissanen 1997, 520)

Yet this usage as well did not eventually survive, and Modern English recruits *you*, *they*, and *people* in the function once served by its birthright *man*.

Rissanen surmises that the disappearance of *me* was due to two factors. If I read him correctly, his proposition that *me* "was too weak for the subject position" leads us to ask why similarly weak forms survive across Germanic, such as the unemphatic forms in Dutch (*je* for *jij*, etc.). For example, to the extent that we analyze the weak forms as kinds of clitics, then we must ask why indefinite *me* could not have simply evolved into one rather than simply disappearing.

Rissanen's other suggestion is that homonymy with the oblique *me*, especially with impersonal verbs (*me semeth*), was a factor. Yet Rissanen himself elsewhere (517) notes that "admittedly, homonymy and disambiguation offer only a shaky argument for the loss of forms," and this is especially apropos here. Crucially, it is likely that in the spoken language oblique *me* and indefinite *me* were not homophones: the former retained a long vowel while the vowel in *me* was likely a weakened one such as schwa (Meier 1953, 179–82, cited in Rissanen 1999). Thus especially given that we are dealing with phonetically similar but hardly identical forms, the homonymy argument is weakened in view of, for example, *det* ([de:] "it" and *de* "they" in Swedish, which, like oblique and indefinite *me* in Middle English, occur with the same verbal ending in the present (since Swedish has but one across person and number). While one might argue that the increasing prevalence of *dom* as "they" in Swedish responds to this homonymy, note that *dej* ([dɛj]) is now established in colloquial Swedish in the second-person singular, creating yet another near-homonymy with *det* even in the absence of conjugational allomorphs of verbal inflection to signal a distinction in meaning. Then beyond Germanic, there are of course cases like *lei* used both as "she" and as a term of polite address in Italian (both with the same third-person verbal ending), and the absence of a number distinction in third-person pronouns in spoken French (*il/ils*, *elle/elles*), despite there being no verb endings to distinguish them.

And meanwhile, obviously neither of these two explanations would apply to the more robust form *man* that persisted alongside its more deeply grammaticalized descendant *me*. One might decide that the disappearance of the distinct indefinite pronoun “just happened.” Yet as Strang (1970, 267) puts it, “No satisfactory all-purpose substitute for it has ever been found,” thus leaving the absence of a single indefinite pronoun to combine with the concurrent disappearance of *thou* in rendering the English pronominal array the most context-dependent of any Germanic language. Of course, in larger view, this is but one of the many developments we have seen that are all symptomatic of a clear trend toward lesser complexity—far beyond the realm of inflectional loss.

4. Implications

It must be reiterated that while I do not claim that English is a “simple” language in the cross-linguistic sense, I do claim that English is significantly less complex overall—according to my metric—than its sisters. The contrast I refer to is illustrated by this comparative table.

	G	Du	Y	Fr	S	N	Da	I	Fa	A	OE	E
inherent reflexives												
external possessors												
gender beyond noun												
loss of prefixes												
directional adverbs												
<i>be</i> -perfect												
passive <i>become</i> verb												
V2												
singular <i>you</i>												
indefinite pronoun												

FIGURE 4.1. Losses in English Compared to Other Germanic Languages

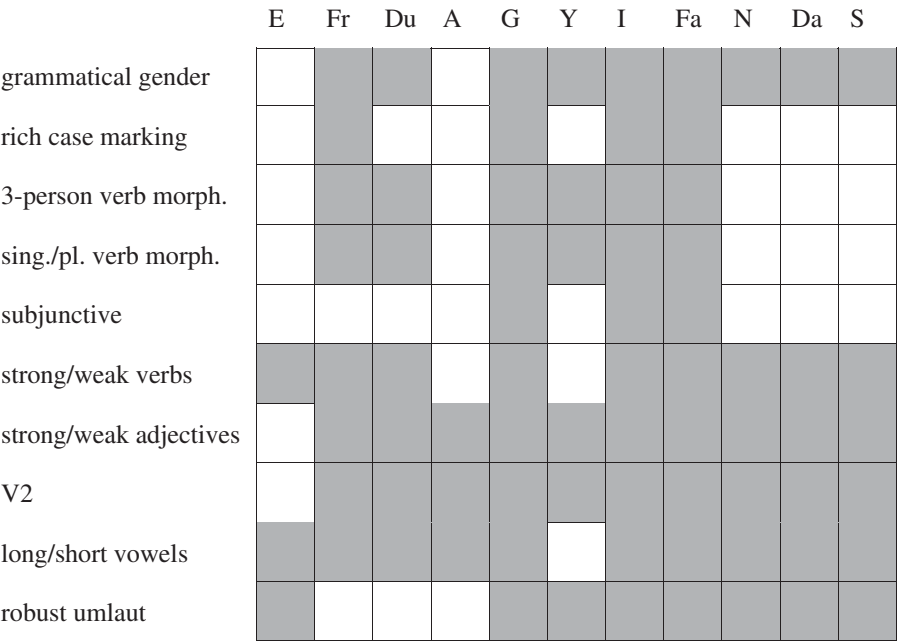


FIGURE 4.2. Losses in English (Lass 1987)

Lass (1987, 318) finds related results in an analogous figure:

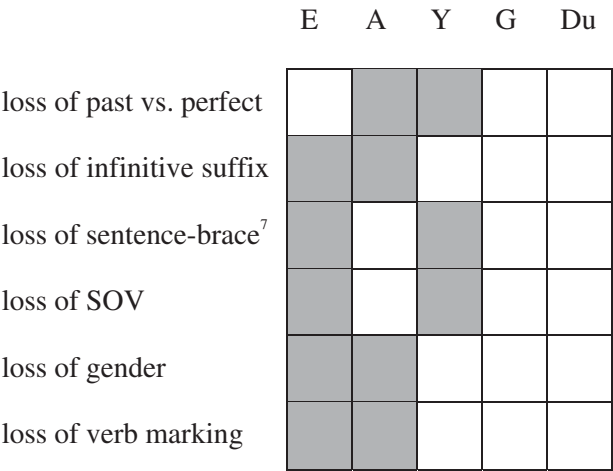


FIGURE 4.3. Losses in the “Contact Heavy” Germanic Languages (Lass 1987)

Lass (ibid., 332) also makes a comparison of English with Afrikaans and Yiddish, specifically, treating these three as uniquely impacted by contact. He again shows that English surpasses the other two in loss (indicated by gray) of Germanic features (German and Dutch are included for contrast):

In Lass's first figure, English and Afrikaans come out even, and are close to it in the second. However, Lass focuses on inflectional features, which automatically brings English and Afrikaans to the fore since both have been shorn of so much inflectional morphology. My treatment, in expanding our perspective to features beyond inflection, demonstrates that Afrikaans is in fact much more conservative than English overall. Importantly, the relative innovativeness of English has consisted not only of the *transformation* of original materials, but of simply *shedding* much of it where none of its sisters have.

Indeed, it is difficult to conceive of any complex of *original Germanic features* that would yield a chart where any Germanic language but English proved to have lost all of the features while the other languages—including English—had retained all or most of them. Obviously, my figure 4.1 can fairly elicit the objection that it is not representative of the Germanic inheritance, my having potentially “stacked the deck” by choosing features absent in English and neglecting the possibility that English may have retained just as many features that many or most Germanic languages have lost.

However, the results are in fact the same even when we bring to bear a more representative array of the Germanic legacy. Below is a generous outlay of grammatical features traditionally treated as tracing to Proto-Germanic, for good measure assembled through reference to two sources, Voyles (1992, 227–79) and Ramat (1998).

Table 4.3. Reconstructed Proto-Germanic Features

- 
1. nominal inflection classes specific to Germanic
  2. nominative, genitive, dative, accusative case marking in NP markers in both numbers on nominals and demonstratives, instrumental in singular
  3. strong/weak distinction in adjectives
  4. masculine, feminine, and neuter adjectival classes
  5. nominative, genitive, dative, and accusative inflections in singular and plural in masculine and feminine, locative, and instrumental inflections in masculine singular
  6. simple singular/plural distinction in neuter
  7. comparative and superlative suffixes *-o*, *-az*
  8. small class of suppletive comparative and superlative forms
  9. interrogative adjectives decline as strong
  10. pronominals in nominative, genitive, dative, and accusative in singular and plural
  11. dual paradigms
  12. reflexive *\*sīn*
  13. inherent reflexive construction
  14. external possessive constructions
  15. one to four declinable
  16. four to ten undecidable
  17. eleven and twelve remnants of duodecimal system (*\*ain-lif* “one left over,” *\*twa-lif*)
  18. thirteen to nineteen composed of unit numeral + “ten”
  19. twenty, thirty, forty, fifty, sixty composed of unit numeral + *\*tigzjuz* “ten-ness”
  20. seventy to ninety composed of genitive plural of unit numeral + derivational *\*-t* + *\*kntom*
  21. 1,000 derived from PIE *\*tūs* “large” + *\*kntom* reflex

TABLE 4.3. (continued)

---

22.	strong/weak distinction in verbs
23.	verbal inflection classes specific to Germanic (class VII forms past and participial form via reduplication)
24.	present and preterite paradigms in active and subjunctive (dual inflections for 1P and 2P except in preterite subjunctive)
25.	present passive indicative and subjunctive paradigms (same inflection allomorph for all numbers in plural in both of the above cases)
26.	infinitive
27.	participles for present, past, and past passive
28.	use of <i>be</i> in intransitive auxiliary + participle constructions
29.	use of <i>become</i> for passive
30.	preterite-present verbs
31.	athematic verbs ( <i>be</i> , <i>do</i> , <i>go</i> , <i>stand</i> )
32.	<i>will</i> -marked “subjunctive”
33.	adverbial suffixes <i>-ē</i> , <i>-ō</i> , <i>-ba</i>
34.	adverbial comparative and superlative suffixes <i>-ōz</i> , <i>-ōst</i>
35.	suppletive <i>well</i> for <i>good</i>
36.	directional adverbs for “at X,” “to X,” “from X”
37.	prepositions doubling as preverbs
38.	conjunctions <i>*endi/*undi</i> “and,” <i>*auk</i> “also,” <i>þauh</i> “but,” <i>*iþa/*uþa</i> “if”
39.	OV word order
40.	large number of derivational prefixes and suffixes

---

It is clear from this list, especially to Germanicists, that it would quite impossible to derive from it a table of features that English has retained *that all or most of its sisters has not*. More to the point, there is not a single feature on the list of this sort. Admittedly, I added a few features from figure 4.1 not traditionally examined in studies of Proto-Germanic: inherent reflexives, external possessor marking, the *be*-perfect, and the *become* passive. However, I am unaware of any other features omitted in Proto-Germanic surveys that are retained in English but rare to nonexistent elsewhere in Germanic, and have the distinct impression that few could identify many—if any.

Certainly English has developed *individual* features of overspecification and structural elaboration that its sisters have not. My thesis is that there was a significant disruption in the transmission of English at one point in its history, but this scenario nothing less than requires that after this, English would naturally drift into its own elaborations as all languages do. Thus English is unique amid Germanic in its *do*-support; in its conventionalization of the present participle with *be* as obligatory in marking imperfectivity in the present, thereby rendering the bare verb zero-marked for habituality; in its distinction of shades of futurity with *will*, *going to*, and *be* + present participle. (It is germane, however, that *do*-support and the development of the *-ing* participle have been compellingly argued to be inheritances from Celtic [cf. 5.1.3] rather than internal innovations.)

Nor, however, have the other Germanic languages simply retained more of the Germanic legacy while otherwise developing no new features. On the contrary, just as we would expect, they developed overspecifications and structural elaborations of their own after branching off from Proto-Germanic. The result was the many well-

known Germanic features absent in English that cannot be treated as Proto-Germanic, such as the conventionalized modal particles in German, Dutch, Frisian, and Mainland Scandinavian; tone in Swedish; and lesser-known cases such as noun incorporation in Frisian.

Thus to wit: while English consists of a massively abbreviated Germanic legacy plus a few later developments, its sisters retain much more of the Germanic legacy *plus* later developments of their own. A claim that English must necessarily be equal in complexity to its sisters is, in the strict sense, illogical. It would require either that 1) *not one* of the several other Germanic languages has drifted into as many new developments as English over the past several centuries, or that 2) English for some reason was uniquely innovative, as if once “burned” by extensive second language acquisition, a grammar is somehow inherently driven to restore a particular degree of needless elaboration. Obviously, neither scenario has any theoretical motivation, leaving the conclusion that English is, indeed, the least complex language in the Germanic group.

## 5. Reassessing the Scandinavian Impact

One response to the contrast in figure 4.1 might be to suppose that English just shed these features by chance. Certain members of a given family have long been observed to be more innovative than others. There are often no apparent “reasons” for this, with linguists treating the difference in conservativity as parents might treat personality differences among their children—they just “are.” However, as I noted in chapter 1, section 3.4, there is a difference between mere statistical flutters in complexity between sisters and starker ones that suggest something at work besides serendipity. English is the latter kind of example.

We will reject two other possible explanations. It is unlikely that English’s departure from the Germanic template was a function of its isolation on an island—generally, this kind of isolation is associated with relative conservatism, Icelandic and Faroese being the obviously pertinent cases here. Furthermore, it cannot be upheld that the standardization of English was the culprit. While one might propose that koineization between speakers of various dialects of English in the London area led to a streamlining of the language, this would leave the question as to why similar processes did not leave standard French (developed amid contact between dialects brought to the Île-de-France region) or Russian (developed amid contact between several dialects in Moscow), similarly simplified in comparison to their sisters. In any case, most of the features I have covered are defunct or on the ropes long before the 1400s.

### 5.1. *Evaluating the Alternatives*

It would seem, then, that the task facing us is to choose between possible contact explanations. I will argue that the break in transmission was caused by the Scandinavian Viking invaders. However, this argument will first require addressing other accounts that have been proposed.



## 5.5.1. NORMAN FRENCH

I accept the arguments of Thomason and Kaufman (1988) and others that the Norman occupation cannot have caused a significant break in the transmission of English. The invaders were too slight in number and removed from the general population to affect the structure of a language meanwhile spoken by millions of people. I would propose that they could not have had any significant impact on the language beyond the lexical (which would thus encompass derivational affixes).

## 5.5.2. LOW DUTCH

I also find it implausible that Low Dutch varieties<sup>8</sup> imported from 1150 to about 1700 by Flemish immigrants and agents of the Hanseatic League (Viereck 1993; Thomason and Kaufman 1988, 321–25), had any significant impact on English grammar. (The term *Low Dutch* encompasses Low German, which some contemporary observers suggest was processed as “the same language” at the time [Peters 1987, 80]). There is nothing in accounts of the Hanseatic League like Pagel (1983) to suggest that Low German speakers were thick enough on the ground in any one place to influence general speech patterns; on the contrary, such sources indicate that the Hanseatic agents were generally housed in their own quarters of town. Moreover, if Low Dutch speakers’ non-native English had influenced the language, we would also expect that their lexical contribution would extend into the grammatical realm. Yet it did not do so to any significant or conclusive extent (Bense 1939), which would include the *-kin* derivational suffix noted by Thomason and Kaufman (1988, 325). These authors (323) also propose that Low Dutch lent several dialects of English an enclitic object form for *she* and *they*, /əs/, but Voss (1995) is rather compellingly skeptical of this account, as well as others concerning sound changes. Overall, despite Viereck’s (1993) and Danchev’s (1997, 101) useful call to consider that Low Dutch had greater impact on English than traditionally thought, I suspect that in the final analysis, we will find that this impact was no more significant than that of Dutch upon English in the New Amsterdam colony in early America. Large numbers of settlers almost always impact the lexicon of the language they encounter, but their numbers alone do not entail transformation of the grammar.

## 5.5.3. BRYTHONIC CELTIC

In recent years, a number of scholars have revived an earlier school of thought that traces the peculiarity of English amid its sisters to Insular Celtic languages (e.g., Poussa 1990; Tristram 1999; Klemola 2000; Vennemann 2001; White 2002). These authors point to evidence in which the traditional historiographical assumption that the Anglo-Saxon invaders all but exterminated the Celts is vastly exaggerated, and that the Celts would have acquired a non-native variety of Old English and passed it on to future generations. The scholars are well aware that by the time English begins its sharp departure from the Germanic template after the Norman invasion, Celtic languages had likely been long extinct among most English speakers; Vennemann (2001, 356) supposes that Celtic was no longer spoken in the north of England after the late eighth

century at the latest. But they propose that this delay may have been due to a diglossic distinction between the written register and common speech, the latter only committed to paper after the “liberation” of English upon the lifting of the documentary “blackout” during the Norman occupation (cf. Vennemann 2001, 364).

Traditionally, specialists in the history of English have been quite skeptical of Celtic influence. While I was once inclined to concur (e.g., McWhorter 2002), the new Celticist evidence, in my view, leaves no doubt that Insular Celtic had a significant impact upon English grammar. The literature is deeply researched, closely argued, and quite copious, and I suspect its position would already be more mainstream if it had not appeared in such obscure venues. Crucially, however, where the impact of Celtic is most unmistakable, it is in terms not of simplification but of transfer. The arguments that Celtic was the source of *do*-support (Poussa 1990; Tristram 1997; White 2002), the *-ing* gerund and progressive (Mittendorf and Poppe 2000; White 2002), and the Northern Subject Rule in which third-person plural is only marked when the verb and the pronoun—but not a full NP—are adjacent (yielding *They peel and boils them* and *Birds flies*) (Klemola 2000) are most likely to gain wider acceptance over time.

The new Celticists would have it that the contact also included simplification, under which features absent in Celtic would be absent in their English. To be sure, because Celtic is so distantly related within Indo-European to Germanic, it lacks most of the features in figure 4.1, such that this scenario seems plausible on its face. But the problem is that where we have conclusive evidence of simplification originating in a particular region and spreading afterward, it is from the north: that is, inflectional loss and the eclipse of V2. This points to agents of simplification in the north, and between the two peoples upon whom English was imposed, the ones who were restricted to the north were, of course, Scandinavians—the Celts were everywhere. Importantly, Celtic constructions appear to have emerged in various locations rather than having spread from one place. It is undisputed that *do*-support began in the southwest, while the Northern Subject Rule began in the north.

For Celtic to lend English aspects of its grammar but not simplify it to any significant degree would be typical of what often happens over several generations of bilingualism—or language shift with interference, in the terminology of Thomason and Kaufman (1988). In their coverage of cases such as the Uralic influence on Slavic, the Dravidian influence on Indo-Aryan, and the Cushitic influence on Ethio-Semitic, language contact resulted in grammatical mixture, but not simplification to any notable degree: replacement prevailed over abbreviation. Russian, Hindi, and Amharic’s grammars are full of influences from languages outside their families, but would never be analyzed as especially “simplified” compared to their respective sisters Czech, Gujarati, or Tigre.

Finally, the frequent claim that Celtic influence would necessarily have left behind a significant lexical imprint is unfounded. We can acknowledge that traditional counts of a mere couple of dozen words from Celtic beyond place names such as Kastovsky’s (1992, 318–19) underrepresent the facts somewhat: Breeze (e.g., 1993), for example, has pointed to many more. Nevertheless, the fact remains that his cases are mostly defunct items that did not survive past Early Modern, or even Old, English. There are, after all, plenty of Scandinavian words that happened not to sur-

vive—which points up that the vast number of Scandinavian words that survive today are a testament to an even more massive initial impact—as well as to how minimal the Celtic impact was in comparison. It would be difficult to provide an explanation as to why hundreds of core lexical items from Celtic for some reason leached out of the language and left but a trickle behind, while a similar battery of Norse ones held on forever.

Or not. As Thomason and Kaufman show, people gradually adopting a language over generations can leave only minor lexical imprint. Subordinated speakers of indigenous languages may adopt a new language and remodel it on the *subconscious* grammatical level, but refrain from the *conscious* transfer into it of native lexical items (Thomason and Kaufman 1988, 116–18). For example, modern Indo-Aryan languages, despite centuries' deep contact with Dravidian languages, have all but no lexical items from them despite profound grammatical influence (Burrow 1955, 380–86). That there was a profound impact from Uralic on early Russian is plain, and yet there are at best about two dozen mostly marginal words in Russian that trace to Uralic (V. Kiparsky 1969).

The Celts would be an equivalent case, and point up the Scandinavians—who left behind a massive lexical imprint—as the more plausible agents of grammatical *abbreviation*. The Scandinavians came to Britain, encountered a new language, and were immediately stuck with floundering about in a version of English that they—big surprise—decorated with words of their own where they could. Celts, original inhabitants who could go on speaking their native languages among themselves as they always had, regardless of the language the new invaders were gradually pressing into their world, got acquainted with the new tongue more gradually.

Finally, just as there are parallels in what Old English lost and Celtic already lacked, as I shall argue in 5.3, there is also the same kind of evidence between the English losses and Old Norse, which was not as completely grammatically parallel to Old English as sources tend to imply.

Thus, on the subject of Celtic transfers into English, I think White (2002, 163) puts it well:

The coincidence seems too much to be that. If we saw such a thing in the Balkans, we would (and should) have no hesitation in regarding it as an areal feature. The present case should not be regarded as somehow different merely because it occurs in Britain.

However, my reading of the arguments on the subject is that if there had been no Viking invasion, Celtic influence would have been, indeed, like that of the Balkan languages upon one another: grammatical transfers without any especial reduction of resources in terms of complexity as defined in this book.

#### 5.1.4. PROCESS OF ELIMINATION

This leaves us with the Scandinavian invaders, who arrived in England in the late ninth and early tenth centuries; roughly speaking, Danes settled in the northeast area that came to be called the Danelaw while Norwegians later settled in the northwest. The Scandinavians settled among the general population rather than ruling from

afar as an elite as the Norman French did, often marrying Anglo-Saxon women. The massive lexical impact of this contact is hardly in dispute, eliminating one serious problem in the Low Dutch and Celtic cases. In terms of sheer number of words, the Normans of course had a vast impact as well. But they imposed their language mainly “from above,” lending mainly content words, generally hewing toward the formal realm. In contrast, the Scandinavian legacy included content words of even the homeliest nature (*neck, window, knife, skirt, happy*, etc.), and extended to grammatical words such as *they, their, them, though, both, same, against*, and others since lost. This alone indicates a highly intimate contact scenario.

Our question, then, is whether the Scandinavian impact upon English went even deeper. I propose that there are indeed indications that the Scandinavian invasions were responsible for the very decrease in overspecification and complexity that I presented in section 3. In this, I will attempt an argument similar to that of Poussa (1982) but in more extended fashion. I will also complement O’Neil’s (1978) observation that English is one of *various* Germanic languages whose development was affected by contact: first, in fashioning an argument within the context of language contact studies as they have progressed since he wrote, and second, by exploring why the degree of reduction in English was greater than in any other Germanic language.

### 5.2. *Support for Scandinavian Influence: Timing*

The first piece of evidence pointing specifically to the Scandinavian impact is evidence that many of these features persisted longest in regions where Scandinavians did not settle, or in those where place-names suggest that they were less robustly represented. Under this perspective, we might assume that these features would be represented even more vividly in these regions even today if a particular dialect that emerged in the Danelaw—standard English—had not gone on to dilute the regional dialects via sociological pressure to such a degree. To put a point on it, if England had remained a preindustrial society where literacy was largely limited to elites, then we might hypothesize that English varieties outside of the Danelaw would remain “card-carrying” Germanic descendants.

The evidence here is solely a first plank in my argument. The vast majority of our substantial sources on regional grammars (as opposed to lexicons and phonologies) was written after standard English came to prevail. As such, in the technical sense, most of our views of nonstandard dialects of England treat varieties in decline, having long ago taken their place on a pole of variation between standard and non-standard forms. However, even these sources give some support to my thesis.

a. *External possessor marking.* Upton et al. (1994, 488) record *wring the neck of him*—rather analogous to the locative Scandinavian configuration of this strategy—in Derbyshire and, pointedly, Cornwall, where there was no Scandinavian settlement at all. (It may also be relevant that they record *He’s pulling that chap his leg* in Yorkshire, the western region of which was rather thinly settled by Scandinavians.)

b. *Grammatical gender.* Scandinavian settlement was concentrated in the north-east; Lass (1992, 113) notes that “loss of inflection [in the noun phrase] is earliest in

the east and north, the south and west generally remaining more conservative." The loss of gender on the definite article began in the north and then was attested only variably in the southwest Midlands in the late 1200s (Lass 1992, 113); meanwhile, the old three-way distinction persisted in the south at this time (Strang 1970, 267), and traces of gender marking hung on in Kent as late as 1340 (*ibid.*, Lass 1992, 113).

Indicatively, of the regions where Upton et al. (1994, 486–87) record the use of *he/him* and/or *she/her* to refer to objects, sixteen out of twenty such usages are outside of Scandinavian concentration (most south of the Danelaw).

There is also evidence that remnants of the gender distinction persisted in especially grammaticalized form in the Viking-free southwest. In the early 1200s, there was occasional gender marking of inanimates there in documents (Strang 1970, 265), and this was still attested in the late nineteenth century in William Barnes's description of Dorset dialect. Here, Barnes (1886, 17–18) describes precisely what we might expect to have evolved in English short of "intervention": a distinction between a common and neuter genders, which he terms "personal" and "impersonal." The "personal" class includes "full shapen things, or things to which the Almighty or man has given a shape for an end" and includes people, living things, and tools: thus of a tree one said *He's a-cut down*, but for water, *It's a-dried up*. The distinction extended to demonstratives ("*theäse*" vs. *this*, "*thik*" vs. *that*).

In reference to inflection in general, Upton et al. (1994, 490) record the infinitive marker *-(e)n* in Derbyshire, Westmorland, and Lancashire, areas of thin Scandinavian settlement (with the exception of *putten*, which occurs in Yorkshire and Lincolnshire as well). They record infinitives ending in [i] only in Monmouthshire, Kent, Cornwall, Somerset, Devon, Lancashire, Derbyshire, Shropshire, and Staffordshire.<sup>9</sup>

In sum, was it merely an accident that the loss of grammatical gender did not begin in the south? The evidence in this paper (including that to come) suggests that it was not.

c. *Directional adverbs.* Upton et al. (*ibid.*, 92) find variations on *come hither* only in regions south of the Danelaw except Lincolnshire. They find *Where to is it?* or *Where is it to?* only in Monmouthshire, Somerset, Wiltshire, Cornwall, Devon, and Dorset (*ibid.*, 502).

d. V2. For one, V2 persists in Kentish documents while eroding elsewhere in English (Kroch and Taylor 1997, 312). These authors' analysis even offers more fine-grained evidence that the erosion of V2 was caused by Scandinavian. They argue (318–20) that the transition from "symmetric" (verb movement in both matrix and subordinate clauses) to asymmetric (verb movement only in the matrix clause) V2 was one occasioned by the inflectional loss that Scandinavian settlers' incomplete acquisition of English led to. Under the assumption that complete loss of V2 would only be possible with sharp diminution of inflection, the implication of Kroch and Taylor's analysis is that the transition to asymmetric V2 was an intermediate stage between the original configuration and today's. By extension, this implies that the disappearance of V2 was initiated by language contact in the Danelaw.

e. *Inherent reflexive marking.* Upton, Parry, and Widdowson (1994, 488) show *sit thee down* and variants persisting in nonstandard dialects throughout England. While they find *laid him down* and *laid her down* only in regions where Scandinavian settlement was relatively thin according to Wakelin (1972, 20)—Cheshire, Derbyshire, and Staffordshire—it is also documented in heavily Scandinavianized Cumberland (Brilioth 1913, 107) and West Yorkshire (Wright 1892, 120; Hedevid 1967, 242). However, it is perhaps notable that Upton et al. only find attestations with other verbs (*ibid.*, 488–89) (*they play(en) them [they disport themselves]*) in Lancashire and Derbyshire. Possibly the usage conventionalized as an archaism with heavily used *sit* and *lie* throughout England (cf. Hedevid’s description of the feature as used only with “certain verbs,” giving the usual *sit* and *lie* as examples [*ibid.*, 242]), but persisted more robustly in areas with less Scandinavian influence.

### 5.3. *Support for Scandinavian Influence: Transfer*

Modern dialectal remnants, however, are only one indication that Scandinavian contact profoundly affected the course of English’s evolution. Transfer evidence provides further, and more striking, support for the hypothesis.

Thomason and Kaufman’s (1988, 302–3) verdict on the evidence of transfer from Scandinavian in English is the following:

The Norse influence on English was pervasive, in the sense that its results are found in all parts of the language; but it was not deep, except in the lexicon. Norse influence could not have modified the basic typology of English because the two were highly similar in the first place.

This conclusion is justified for Thomason and Kaufman’s masterful argument on the basis of the features they treat as part of their “Norsification package.” However, other evidence suggests that the Norse influence was indeed deep, and that Thomason and Kaufman’s Norsification package, comprising mostly phonological and morphological traits, various grammatical items, and some lexical items, constitutes but a subset of the relevant evidence.

According to tradition, Thomason and Kaufman assume that Old English and Old Norse were too similar for structural transfer to be particularly relevant to analyzing their effect upon each other in contact. Yet a closer look at Old Norse reveals grammatical differences crucial to this thesis. Specifically, no fewer than six of our ten losses in English have parallels in aspects of Old Norse hitherto overlooked, to my knowledge, in studies of the Danelaw situation.

a. *Inherent reflexives.* In Old Norse, the reflexive use of the first-person and third-person pronouns had eroded and affixed to the verb as a suffix, the latter used in all persons but the first singular: *bindomk* “I tie myself,” *bysk* “you arm yourself,” *staksk* “he stabbed himself” (Heusler 1950, 107). This extended to inherent reflexives: *þeir setiask niþr* “they sat down,” *er hefnezk á honom* “you revenge yourselves upon him,” *þetta felsk honom vel í skap* “that felt good to him, agreed with him” (*ibid.*, 137–38). Use of free pronouns in the reflexive was not unknown, but was largely restricted to

dative forms (*hann brá sér* “he wandered”), but even here was variable (*hann brásk* was also grammatical) (ibid., 138).

The variability of inherent reflexive marking by Middle English may have been the result of a tendency for Old Norse speakers to omit the inherent reflexive pronouns in speaking English. Use of the full pronoun was the marked case in their native language, and, meanwhile, English lacked any equivalent of Old Norse’s reflexive inflection. English specialists often note that morphology could be shed in the Danelaw because it was incidental to communication. Morphological marking of inherent reflexivity would have fallen under this rubric by definition.

Thus because it was encoded morphologically, Old Norse speakers would have been comfortable refraining from marking the distinction when speaking English—just as Modern English speakers are.

b. *External possessives.* Scandinavian is unique in Germanic in encoding external possession with the locative rather than the dative (cf. [7], [8], [9]). Faced with this disjunction between external possessor encoding in the two languages, Scandinavians may have taken the choice of eliminating the distinction altogether, given that it was not vital to the expression of the relevant concepts. This is a common process in the development of koines, for example, where often, the koine eschews features which were present in most or all of the source varieties, but expressed with different morphemes or strategies (cf. discussion in 5.4 below).

c. *Derivational prefixes.* The absence of the core Germanic verbal prefixes in Icelandic and Faroese traces back to Old Norse (Heusler 1950, 40). It could be that the rapid eclipse of these prefixes in English was due to the absence of cognates in Old Norse speakers’ native language.

d. *Be-perfect.* In Old Norse, as in Modern Icelandic, the *be*-perfect largely connoted the resultative and the passive (Heusler 1950, 136). Its use as a true perfect was limited to a few intransitive verbs such as “to go”: *ek em gengenn* “I have gone” (Brenner 1882, 129). This may have been the spur for the disappearance of the feature in English, including the possibility Bruce Mitchell (1985, 302–4) notes that what has been analyzed as a *be*-perfect in Old English may have actually been *only* a resultative construction.<sup>10</sup>

e. *Become-passive.* In Old Norse, the passive was usually expressed with *vera* “to be.” *Werða* “to become” was relatively restricted in meaning, encoding roughly the saliently active semantics of Modern English’s *get*-passive:

(20) þær saker skal fyrst dōma,  
the.PL issue.PL should first adjudge.INF

er fyrra sumar varþ eige um dōmt.  
that last summer become not to judgment

‘First the complaints should be decided upon that didn’t manage to get to judgment last summer.’ (Heusler 1950, 137; translation mine)<sup>11</sup>



This restricted usage may have been a cause of the otherwise mysterious absence of English's *weorðan* after Old English.

f. *Indefinite man*. Icelandic and Faroese lack a *man*-cognate, and already in Old Norse it was recessive (Heusler 1950, 147), generally replaced by impersonal verb constructions or third-person verbs without pronouns. This may possibly have set in motion a de-emphasis on the use of Old English's *man*-cognate that eventually resulted in its disappearance early in Middle English.

English is traditionally considered closest to Dutch and Frisian. But in many aspects where Old Norse and its descendants depart from the Germanic pattern, they parallel English, although having overall retained a great deal more of their Germanic legacy. Only Old Norse and its modern descendants offer anything approaching six out of ten features which could be expected to result in elimination in a contact situation with English, a clustering which the author did not even expect to find upon beginning this investigation. This—especially in congruence with the dialectal evidence adduced in 5—suggests a specific effect from Scandinavian. In addition, it further weakens the potential import of the correspondences between English and Celtic, given that Old Norse offered just as many features which we could expect to occasion loss in English in a contact situation—but *in addition* left behind a rich lexical legacy.

To wit, in reference to Thomason and Kaufman's statement that "Norse influence could not have modified the basic typology of English because the two were highly similar in the first place," the similarity dilutes somewhat on closer examination, to an extent that may well have had significant impact on the development of Modern English.

#### 5.4. *Support for Scandinavian Influence: Reduction*

Generally, discussion of the Scandinavian impact on English is largely restricted to transfer effects: sound changes and lexical borrowings. I have attempted to add possible structural transfers to the relevant discussion. As such, however, it might be objected that my attempt to expand our conception of the Scandinavian impact is hindered where Old English and Old Norse have parallel structures, under the assumption that English would likely have retained these features rather than shed them.

But as Thomason and Kaufman (1988, 129 *passim*) note, when a population shifts to a new language and their rendition of the language ousts the original native one, then transfer effects often occur alongside evidence of incomplete acquisition—that is, outright *reduction* occurs rather than transfer. Importantly, these effects often occur even where the languages in question have parallel or cognate structures.

##### 5.4.1. TRADITIONALLY ACCEPTED: INFLECTION AND GRAMMATICAL MORPHEMES

Scholars of English's history traditionally recognize this in the area of inflection, where it is often reconstructed that when Old English and Old Norse speakers were con-



fronted with equivalent but phonetically differing inflections, they simply shed them to ease communication (O'Neil 1978, 256–60, being an extended presentation). Yet for our purposes we must refine the conception somewhat. Most writers appear to suppose that this must have been a two-way affair, communication occurring as speakers of both languages shaved away their native inflections, leaving an analytic “common denominator” comprehensible to both.

However, it is questionable that this would have allowed significant communication. Lass (1987, 52) and Kastovsky (1992, 328–29) suggest that claims that Old English and Old Norse were virtually mutually intelligible are exaggerated, and I would agree—this idea strikes me as reflecting the written medium, allowing us to “wrap our heads around” phonetic correspondences at leisure, and concealing the obstacles to comprehension effected by accent, intonation, and morphophonemics. Analyzed side by side, Old English and Old Norse suggest no more mutual intelligibility than that between, for example, Serbo-Croatian and Bulgarian or Spanish and Italian.

Moreover, mere contact between closely related varieties hardly *entails* the dilution of infection. Contact alone, amid extensive *first-language* bilingualism, can lead to a language with a mixture of inflections from several languages, but no less inflected overall than its precontact form. *Pace* arguments like O'Neil's (1978, 256–60), are even true when languages are closely related, despite initial impressions that the confusions created by cognate affixes would simply drive speakers to shed morphology altogether. One example is Rusyn, formed through contact between dialects of Ukrainian, Slovak, and Serbo-Croatian (Harasowska 1999).

Inflectional *loss* is symptomatic of a more specific phenomenon: non-native acquisition. Specifically it is more likely that the inflectional loss resulted from Old Norse speakers' incomplete acquisition of English (cf. Danchev 1997, 90). Thomason and Kaufman's (1988, 119–20) stipulation that interference effects are likely to be stronger in cases of rapid shift over “one or two generations” is also relevant, given that these authors are of the opinion that Norse was no longer spoken in England after 1100 (*ibid.*, 282), disappearing within one or two generations of the reintegration of the Danelaw into the English polity (*ibid.*, 284, 286).

The Norse grammatical morphemes that English incorporated make a further case for this. A language generally only borrows grammatical, as opposed to lexical, items in contexts of bilingualism (amid at least a subset of the population). The question here is whether the Norse were more likely to be bilingual—thus developing an English bedecked with Norse grammatical items—than the native English speakers were.

There would appear to be little question here. The Danes and Norwegians were newcomers, who were largely illiterate and thus did not impose their language in writing or in government, and eventually gave it up. Obviously the impulse toward bilingualism would have been much stronger among the Vikings than among the English. In this light, items such as *they/them/their*, *both*, *same*, and so on would stand as remnants of Scandinavian brought into the English spoken by, first, immigrants, and then just as plausibly by succeeding generations bilingual in Old Norse and Old English. It is rarely if ever acknowledged that after the first generation, descendants of the Scandinavian invaders may well have begun to speak English as well

as Norse even among themselves, as is typical of shifting speakers, with Norse and English perhaps taking their place in a kind of diglossia. As such, for Norse descendants born in England, an English sprinkled with the occasional Norse grammatical item would have been not only comprehensible, but even a marker of, if we may, “ethnic” kinship.

#### 5.4.2. EXTENDING THE PARADIGM

These observations, then, suggest that English could easily have shed even features it *shared* with Old Norse, if its fate was determined by non-native speakers’ proficient yet approximate rendition, as the timing and transfer evidence suggest. This is in itself hardly a venturesome proposal, the conception having long been considered unexceptionable as applied to morphology, perhaps because the difficulty of acquiring a foreign language’s morphology is so familiar and readily perceptible. My suggestion is simply that we extend this mechanism beyond morphology, especially given that modern language contact studies offer no grounds for supposing that this would be scientifically inappropriate.

The relevant comparison, for example, is with koines, such as Siegel’s (1987, 185–210) description of the koine Hindustani of Fiji, developed amid contact between speakers of a range of divergent varieties of the Hindustani dialect complex. Certainly, to an extent the koine has picked lexical and morphological features “cafeteria-style” from assorted dialects rather than shedding them. However, on balance, the koine is not as elaborated as any of these dialects, instead being markedly simpler in the formal sense than any of these, *even when the dialects all display the feature in question*. This includes the elimination of the three-way formality distinction in second-person pronouns (ibid., 199), a general tendency toward replacing synthetic with analytic forms (ibid.), and a strong tendency to replace SOV order with SVO (ibid., 198). (This last is especially indicative regarding the loss of V2 in English.)

A less well-known example is the Riau dialect of Indonesian (Malay) described by David Gil (1994). While developed amid speakers of languages closely related to Indonesian and to one another, Riau Indonesian has vastly simplified Indonesian’s valence-marking morphological apparatus—of much greater semantic and syntactic import than mere gender or person/number markers—and other grammatical features with close cognates in the languages spoken by its creators, such as Minangkabau. These cases demonstrate a phenomenon well documented by second-language specialists, in which interlanguages can easily lack features present in both L1 and L2 (cf. Pienemann et al. forthcoming; Håkansson, Pienemann, and Sayehli 2002; Habertzettl 2005)—an especially useful example for this chapter is Håkansson et al.’s observation that Swedes’ interlanguage when learning German lacks V2 although both Swedish and German have it.

These cases demonstrate that even when languages in contact are closely related, reduction can play as significant a part in the outcome as exchange of materials—*far beyond mere inflection*. In processes of linguistic accommodation, speakers often contribute a less overspecified and complexified rendition of their language. The extreme manifestation here is Foreigner Talk; a less radical manifestation would be the tendency for creators of creoles to contribute a “streamlined” version of their native

grammars to the new language, such as among Oceanic speakers (Keesing 1988; Siegel, Sandeman, and Corne 2000). Koine scenarios exhibit an analogous process, and there is no theoretical reason that this would not have been the case in the Danelaw.

In both the Fiji and Riau cases, the scenario is complicated somewhat by the possible impact of pidgin varieties in the contact situations: here, Pidgin Hindustani and Bazaar Malay, respectively. However, this may be an artifact of our temporal proximity to cases such as these. It is hardly inconceivable that there was a “pidgin English” spoken by the first wave of Scandinavians, reflecting the limitations of adult language-learning capabilities. (This is especially the case given the observed fact that Old English and Old Norse were not essentially dialects of the same language as is often claimed.) It may have been the progeny of these invaders who acquired a more proficient English, nevertheless pervaded with Norse features. Our microsociolinguistic knowledge of how interference through language shift (in Thomason and Kaufman’s terminology) proceeds is currently limited, most cases having occurred beyond the purview of written history. Certainly this is, and will likely remain, the case with the Norsification of English over a thousand years ago.

#### 5.4.3. TWO MECHANISMS FOR FOSTERING UNDERSPECIFICATION AND SIMPLIFICATION IN ENGLISH-NORSE CONTACT

1. *Trigger weakening.* In many cases, already in Old English features were ripe for marginalization in a contact situation, because they occurred only variably. This is the case with inherent reflexives, external possessors, directional adverbs, the *be*-perfect, the *become* passive, V<sub>2</sub>, and indefinite *man*.

It is tempting to hypothesize that this very optionality of the features in Old English was due to Scandinavian contact. But this is unlikely given that most Old English documentation from 900 AD onward is in the West Saxon dialect, outside of Scandinavian settlement. Crucially, however, this hardly means that these features were used in more tightly conventionalized fashion beyond West Saxon. On the contrary, this kind of variability is attributable to speech varieties that are primarily spoken, as was true of Old English—while it comes down to us in writing, it was spoken in a society where literacy was largely limited to an elite. The especial prevalence of “unfocused” conventions is familiar to any linguist working on an unwritten indigenous language, and even characteristic of other early Germanic varieties like Gothic and Old Norse itself (in the latter of which inherent reflexives, *become* with the passive, and indefinite *man* were variable). As such, we can assume that the “softness” of the relevant rules was typical of English dialects in general, not just West Saxon. (This also discourages supposing that the variability of many features in Old English was due to Celtic influence “softening up” the grammar initially, with Scandinavian contact merely reinforcing a process already begun. If other early Germanic languages unaffected by Celtic already displayed similar variability, then the Celtic explanation for English loses necessity.)

In any case, this would mean that such features were especially vulnerable to falling below the line of acquirability, via a process outlined by Lightfoot (1997, 1999).

Namely, the general tendency toward isomorphism and simplification in the generation of non-native varieties would have led Norse speakers to use these features even less in their rendition of English. This could have initiated a decline in frequency of occurrence over generations of the sort that Lightfoot analyzes as weakening and eliminating the sufficient “trigger,” in his terms, for its transmission to new generations. Importantly, the weakening of many of these features would have been reinforced by being variable or marginal in Old Norse itself: inherent reflexives, the *be*-perfect, the *become*-passive, and the indefinite *man*.

A modern parallel would be a feature of the Irish Gaelic spoken by second-language learners in Ireland (Henry and Tangney 1999). Equative predications in Irish Gaelic mark the subject as accusative, place it clause-finally, and use a distinct *be*-verb (*Is dochtuir é* [be doctor him] “He is a doctor”). This, however, is an exceptional feature in a language in which “ergative” constructions are otherwise nonexistent. English-dominant learners tend to replace this quirk with a construction using the more commonly encountered *be*-verb *tá*, which occurs sentence-initially as verbs typically do in this VSO language, and takes a nominative subject: *Tá se dochtuir* [be he doctor].

Importantly, Lightfoot’s framework entails no stipulation that a given feature would vanish immediately. Rather, once initiated, the weakening of a trigger proceeds gradually toward its total obliteration, each generation using a feature less often than the previous one and providing even less stimulus for the next one. Lightfoot’s work indeed has emphasized “catastrophic” parameter shifts. However, he makes no stipulation that the process of trigger weakening that eventually causes these parameter shifts is, itself, of such an abrupt nature (and to the extent that his earlier work claimed that such changes occurred in highly narrow time frames; in later work he has considerably tempered this perspective). As such, my claim is not that contact in the Danelaw would have obliterated the features in figure 4.1 immediately, a scenario obviously impossible to support given the documentary evidence.

Certainly the optimally “clean” version of my argument would demonstrate that all of these features evanesce from English before 1200 AD—but the documentary evidence obviously does not support such a scenario. Our goal is to construct a scientifically responsible explanation to account for the *gradual* disappearance of the features in question—this being as imperative to explain as their more abrupt disappearance would be, given that they were retained in a dozen-odd sister languages. Lightfoot’s framework provides a basis for proposing that Scandinavian contact “tripped off” a decline of these features, that only later culminated in complete disappearance.

2. *General “trimming” of overspecified and complex features.* Other features may have been eliminated even when robust in both languages. To reiterate, this is the very conception that is common consensus regarding grammatical gender in the noun phrase. I suggest that it was also the case with the collapse of gender distinctions on the articles. A non-native speaker of English, confronted with three forms of the article corresponding to gender assignments that often conflicted between Old English and Old Norse, would plausibly have made do with a single gender-neutral marker rather than applying native genders to his or her version of English. Koine data worldwide, such as Fiji Hindustani discussed above, indicate that when structures are

cognate but distinct in closely related varieties, the speaker of one of them is as likely to eliminate her or his reflex of that feature as to preserve it—even when doing the latter would not appreciably impede communication.

I also suggest that the gradual disappearance of V2 falls under the same rubric. An analogy is the weakness of SOV order in Fiji's Hindustani koine, created by people who spoke languages where SOV was obligatory. Here, the issue was less structural discrepancies—minor in both the Danelaw and Fiji—than the simple impulse toward simplification in non-native speech varieties.

#### 5.4.4. WHERE ART *THOU*?

Of the ten features in figure 4.1, only *thou* is incommensurate in its history with a thesis that conditions in the Danelaw in general were the key to English's detour from the Germanic template. *Thou* disappears only in the written standard (and even then rather late), but remains robust in nonstandard dialects even within the Danelaw region, regularly appearing in dialect surveys even of twentieth-century speech.

One might suppose that the demise of *thou* in one dialect was merely an accident, just happening to occur in the region that went on to produce the written standard. However, the uniqueness of the development compared with (all dialects of) all of English's sisters—as well as with European languages in general—remains striking. As a chance occurrence, the eclipse sits as an eternal loose screw—but a more elegant analysis subsumes it under the general trend in English's development toward underspecification. To suppose that the death of a distinct second-person singular pronoun was mere happenstance means that one would find it unremarkable to see the decline occur in a hypothetical English where the other features of figure 4.1 remained intact. In light of the fact that not a single one of the other Germanic languages has taken this route over millennia of existence, I propose that such a development would have been highly unlikely.

Just why this occurred only in the East Midlands dialect will most likely remain a mystery. Perhaps it lends support to Poussa's (1982) argument that this variety was a particularly koineized one serving as a lingua franca between speakers of still-divergent Middle English dialects. But our perspective on *thou* must be informed by its utter anomaly, implausibly linked to any cultural factor local to England, and impossible to trace to the influence of Latin on English standardization, since Classical Latin evidenced no such trait. In being the only Germanic language refraining from distinguishing number in the second person—a cross-linguistically rare trait that the typologist associates with obscure, isolated cases such as the South American language Pirahã and a few languages in Papua New Guinea and Indonesia—standard English displays one of a great many underspecifications which, in their sheer number, suggest an external explanation.

#### 5.5. *The Nature of Evidence*

Obviously at this point we yearn for particulars of the sociolinguistic terrain in the Danelaw. However, this was a largely illiterate setting of primarily oral communication lost to the ages; the writing that has come down to us was, in broad view, a

marginal activity aimed at a small elite. Even the numerical size of the Scandinavian presence remains controversial—and most likely unknowable. Thomason and Kaufman (1988, 276, 361–63) cite Sawyer's (1971) arguments that there is no evidence that vast streams of peasant settlers followed in the wake of the invaders as often supposed.

However, I propose that the clustering of the changes in the Danelaw region where Scandinavian settlement was heaviest, combined with the correspondences between the losses in English and features absent or marked in Old Norse, implicate the Scandinavian invasions *despite* the lacunae in the sociohistorical record. Sawyer may be correct that this impact was due more to prestige than brute numbers, and the questions here naturally evoke an interest in the precise mechanics of how a language is transformed by widespread second-language acquisition. But at the end of the day, a preliterate society 1,200 years removed from us is unlikely to ever shed much light on this question.

To serve our interest in the on-the-ground aspects of this kind of process, there are various living or amply recorded analogues to what I propose happened in the Danelaw (e.g., see LaCroix 1967 on vehicular Fula, or Siegel 1987 on pidginized and koineized Hindustani). My aim has been to demonstrate that even if we will likely never know precisely *how* the transmission of English was temporarily diluted in northeastern England, there is evidence allowing us to know that this *did* happen.

Clearly, my hypothesis can be vastly strengthened with closer engagement with the document literature, computer corpus-based techniques being particularly germane. However, I propose that such engagement would be a worthwhile pursuit, as it offers a possible answer to a contrast which must otherwise stand as a pressing anomaly—to wit, figure 4.1.

## 6. English versus Afrikaans

A perhaps surprising conclusion of this thesis is that English experienced a significantly greater disruption of transmission during its timeline than Afrikaans, despite the now established analysis of Afrikaans as a “semicreole.” Viewed in comparison to its Germanic sisters through the lens of figure 4.1, Afrikaans differs largely in its loss of inflections, with the mere addition of loss of the external possessive construction and indefinite *man*. Overall it remains very much a “well-behaved” Germanic language. The losses in English vastly surpass this degree.

That the difference between English and Afrikaans indeed suggests a difference in *degree* of transmission disruption is suggested by evidence that inflectional loss is a “first layer” along the cline toward outright pidginization. Along the cline toward pidginization follow losses in other grammatical features characterizable as “ornamental” upon the language competence (cf. McWhorter 2001). For example, there are contact languages that lack most or all of the inflections of their source languages but largely retain the remainder of their source languages’ grammatical machineries, but none that lack source language features such as derivational markers, inherent reflexive marking, evidential markers, or inalienable possessive marking while retaining source language inflections.

Of course, nevertheless, English also displays internal developments departing from the Germanic template which, in their “elaborative” nature, discourage any analysis of the language as “creole” or even “semicreole”: *do*-support, and so forth. However, if my thesis is valid, then the reason English contains features that distract modern observers from the degree of its overall anomalousness amid Germanic is that it underwent a profound impact from extensive non-native acquisition so much further in the past than Afrikaans.

Afrikaans emerged from Dutch just after three centuries ago, while the Scandinavian encounter with English occurred over a thousand years ago. After this, English had roughly five centuries to evolve before the invention of printing and the advent of widespread literacy, thus remaining relatively unfettered by the effects of prescriptivism and standardization. Moreover, only in the late eighteenth century would prescriptivism become a truly decisive force upon general tendencies in English speech. As a result, after being “shorn” of so much of its complexity, English had as long as seven hundred years to “get back on its feet,” returning to the tendency toward drift into elaborations that is typical of natural languages as they are spoken over time. Crucially, *all of the internal elaborifications of English adduced above occurred long after the nativization of the Norse-speaking invaders*. Afrikaans, meanwhile, has existed for only a few centuries, and, in addition, within this time the language has been codified as a written language, its natural change thus retarded.

Thus the reader is asked to imagine that printing and widespread literacy had arisen in England in, for example, 1000 AD, the language thus subjected to standardization and prescriptivist impulses shortly after the Scandinavian impact, with internal developments in the nonstandard dialects marginalized or stanchied by the dominance of the standard as is the case today. I suggest that this alternate-universe standard English—roughly early Middle—devoid of *do*-support, *going to* versus *will*, or *I am speaking* versus *I speak*, would stand out just as much as Afrikaans today as a peculiar abbreviation of Proto-Germanic.

Of course, in reality, Gutenberg and universal education arose much later. However, a comparative analysis reveals that even seven hundred years could not erase what appears to have been more far reaching an interference in transmission in English’s history than that of Afrikaans.

## 7. Conclusion

My aim has been to show that the difference between English and its sister languages comprises a much larger array of features than merely loss of inflection or a “tendency toward analyticity,” and that the larger awareness that English has moved toward a “different typology” than its relatives (e.g., Lass 1987, 317–32) is traceable to a causal factor rather than being a random “uninteresting” development.

As many readers may have noticed, my list of features is hardly complete. Radiating outward from the core of losses that leave English unique or close to it in Germanic, there are other losses that English shares with a few Germanic languages, their interest being that where a subset of Germanic languages have departed sharply from the original Germanic “typology,” English never fails to be a member. Examples



include subjunctive marking (lost or marginal in Mainland Scandinavian, Dutch, Frisian, and Afrikaans) and verb-final word order in subordinate clauses (lost in Mainland Scandinavian, Icelandic and Faroese, and Yiddish). English is also alone in Germanic in lacking a strong-weak distinction in adjectival inflection (even Afrikaans retains this based largely on syllable count), which I omitted from my presentation to detract from the traditional focus on inflectional loss in studies of English's history. Its inclusion, however, would obviously only bolster the argument.

Moreover, it is certainly true that the Mainland Scandinavian languages shed a great deal of inflection. However, I have shown that beyond inflection, these languages retain strikingly more of the Proto-Germanic inheritance than English. In addition, Thomason and Kaufman's implication (1988, 315–18) that even the inflectional loss was simply a matter of chance does not fare well in the face of other research. Many specialists readily attribute this loss to non-native acquisition just as I attribute the English losses; specifically, by speakers of Low German (Haugen 1976, 313; Jahr 2001; Kusters 2003, 192–93). Overall, the Germanic languages fall upon a cline of reduction, roughly the following: Icelandic/Faroese/High German/Dutch, nonstandard Germans, and Yiddish/Mainland Scandinavian/Afrikaans/English. Subsequent research may tease out the roles, possibly interactive, that contact and fortuitous internal elisions played in the histories of these languages to create this cline. For our purposes, the crucial issue is that contact had such an unusual effect in the English case.

I must reiterate that my claim is neither that English is a “creole” nor that elaborations have been alien to its life cycle. However, our full cognizance of *do*-support, and so forth must be seen in conjunction with the striking losses that English suffered in the centuries during and in the wake of the Scandinavian invasions, of a disproportionate volume leaving it much more underspecified or complexified *overall* than even Afrikaans.

And the result is that English is the most context-dependent grammar in the Germanic family, by a strikingly wide margin. Of course, if there were only three or four Germanic languages, then serendipity might remain a plausible explanation of the discrepancy. But instead there are about a dozen Germanic languages, each of them subsuming a number of dialects. To assume that the nature of the changes in English was merely a matter of chance would seem to require that the slings and arrows of outrageous fortune had led at least one of the many other Germanic languages—or even dialects—to shed, rather than transform, a comparable volume of the Germanic legacy.

Otherwise, the conclusion would seem almost unavoidable that the English timeline was decisively influenced by what Trudgill (2001) has termed, in apt and savory fashion, “the lousy language-learning abilities of the human adult.”



# Mandarin Chinese

## *“Altaicization” or Simplification?*

### 1. Introduction

Linguists are familiar with the linguistics “factoid” that Mandarin Chinese has four lexically contrastive tones while Cantonese has nine. However, it has been less remarked upon that this difference is but one symptom of a distinctly lesser degree of elaborification in Mandarin than in not only Cantonese but also in all of the other Chinese languages. The linguistic and historical evidence strongly suggest that like English, Mandarin Chinese was profoundly impacted by extensive second-language acquisition during its life cycle.

Hashimoto (1986 and elsewhere) has indeed addressed Mandarin’s uniqueness, but within an argument that Mandarin is distinct from the other Chinese varieties in having incorporated transfer from Altaic languages. Hashimoto’s argument has been accepted by specialists, but besides its presentations being rather brief, it has not been processed within an acknowledgment that the difference between Mandarin and its sisters runs much deeper than a difference between Altaic transfer in the north and transfer from Tai and Miao-Yao languages in the south, as Hashimoto, Ballard (1985), and others have argued.

Hashimoto (1986, 77) does suggest that Altaic-speaking learners “pidginized” Mandarin. But he only explicitly exemplifies this in his surmise that Altaic languages’ initial-stress tendency led their speakers to decrease the number of tones in their rendition of Chinese, and in a suggestion that incomplete acquisition decreased the number of classifiers in spoken Mandarin (*ibid.*, 93). His main focus in his 1986 article and elsewhere is upon transfer, such as of word order, or choice of causative morphemes like *ràng* “let” to encode the passive rather than *give* as is typical in southern varieties (Hashimoto 1987). This aspect of his argument has had the most influence on comparative Chinese dialectology, in which an “Altaic transfer in the north and Tai/Miao-Yao transfer in the south” schema has become canonical.

But in this chapter I will demonstrate that while Altaic languages did have a decisive effect upon Mandarin, transfer was a relatively minor factor, while reduction due to nonnative competence was comprehensively transformative. The evidence also suggests that the crucial locus of transformation was not, as is often supposed, the occupations by Genghis Khan (thirteenth through fourteenth centuries)<sup>1</sup> or the Manchus (1644–1911), but the widespread resettlement of conquered and dispossessed peoples amid Han Chinese on the northern Chinese frontier from the 600s through the 800s during native Chinese rule under the Tang dynasty.

## 2. Preliminaries

### 2.1. Mandarin's Internal Developments

Just as English has local overspecifications that its Germanic sisters lack, Mandarin is hardly less overspecified than its sisters without exception. Sheer serendipity would render this unlikely; our argument is that Mandarin is less overspecified overall.

Thus Mandarin has its famous “patient”-marker *bǎ* (with all due acknowledgment of the widely discussed fact that this is a highly oversimplified characterization of its function):

- (1) Házǐ bǎ fàn chī-wán le.  
 child 'ACC' rice eat-finish ASP  
 'The child has eaten the rice.' (Chao 1968, 342)

While cognate markers occur in other varieties, Chappell (2001a, 335–36) and others argue that these usages are based on the prestige of Mandarin, being “written” or in formal constructions not indigenous to the spoken varieties. This marker, then, is a Mandarin innovation rather than a pan-Chinese feature.

Mandarin has other local developments as well. For example, Chappell (2001b, 74–75) shows that Mandarin's experiential aspect marker *guò* has evolved an additional completive phase-marking function that its cognates in other Chinese varieties have not. This kind of local development is what we would expect of a language which has been transforming internally since its birth, and there is no grounds for supposing that this kind of transformation would not continue after an interruption in its time line. Our claim here will be that these emergences in Mandarin leave it nevertheless anomalously less complex *overall* than its sisters.

### 2.2. Register Layerings

It is also well known that the nonstandard Chinese varieties are complexified not only by internal developments or retentions of original materials but also by layerings from Mandarin due to its influence as a literary standard. The phenomenon is especially well studied in Min, with its high/low lexical and morphological doublets reminiscent of the distribution of Latinate/French lexical items and original Germanic ones in English (Sung 1973; Lien 2001), and a similar layering of types of interrogative construction (Yue 1991) (cf. also Matthews and Yip 2001 on relative clause constructions in Cantonese).

However, again our argument is a relative one: We will see that the nonstandard Chinesees would dwarf Mandarin in overspecification and structural elaboration even if Mandarin had not exerted this influence. Moreover, it must also be noted that Mandarin itself displays an analogous sort of register layering due to the prestige of Classical Chinese forms, whose mastery is central to nuanced expression in the standard language (Li and Thompson 1982). The Sinitic sisters are essentially at “par” here, then, while our interest is in differentials in complexity beyond this feature that the languages have in common.

### 2.3. *Gradient Mandarin Influence*

Finally, related to the above facts, Mandarin does not display its uniqueness with the degree of “splendid isolation” as does English in relation to its Germanic sisters. Rather, the most salient contrast in elaborification is between Mandarin (Northern) and the Southern varieties: Hakka, Yue (including Cantonese), and most strikingly Min (including Fujianese on the mainland and Taiwanese). The Central varieties—Wu (including Shanghainese), Gan, and Xiang—are distinctly more elaborified than Mandarin, but to a lesser degree than the Southern languages.<sup>2</sup> The Central varieties have undergone heavy influence from Mandarin as the result of their proximity to Mandarin-speaking regions as well as population migrations through the ages (e.g., LaPolla 2001); the same factors are even considered to have had a lesser but like effect upon Hakka (cf. Dryer 2003, 48–53, on Hakka’s frequent modifier-head order as in Mandarin).

Hashimoto (1986, 76–77) and Chappell (2001a, 336) refer to this gradation in their observation that “Altaic” influence spreads southward from Mandarin—although I would argue, on the basis of this chapter, that what spread was actually more the general reductive tendency in Mandarin than its rather light Altaic inheritance.

In any case, the argument here is not, then, that Mandarin is an “English” uniquely denuded of a comprehensive range of proto-language features still resplendent in all of its sisters. Instead, for one, Mandarin simply tends to display Sinitic features in less complexified form than any of its sisters, and, in addition, the sisters display these features more elaborately in a gradient increase, roughly, as one travels due southeast.

Nevertheless, even the contrast between Mandarin and the Central varieties remains striking enough to beg explanation, while that between Mandarin and, especially, Yue and Min is so vast as to virtually require that the latter represent an uninterrupted development that was at some point broken in Mandarin.

## 3. First Demonstration Case: Mandarin versus Cantonese

As a preliminary demonstration, the type and degree of difference I refer to is particularly well exemplified by a comparison of Mandarin and Cantonese, since the latter is both typical of the high degree of elaborification in southern Chinese as well as being the best studied of the nonstandard varieties.

In light of the previous chapter, roughly speaking, Cantonese is Hochdeutsch to Mandarin's Dutch. Mandarin retains ample amounts of the Old Chinese legacy, and has its occasional internal developments just as Dutch has, for example, pragmatically conventionalized diminutive marking to a unique degree among Germanic languages. But overall, Cantonese displays more overspecification and structural elaboration, including requiring attendance to more semantic distinctions in a phonetic string. Chinese is especially useful in demonstrating that inflectional affixation is but one of a wide array of features relevant to comparative complexity across languages. Given Chinese's analytic typology, the difference between it and its sisters manifests itself much less in affixal features than in free morphemes.

Unless otherwise indicated, the sources for the data in this section are Li and Thompson (1981) for Mandarin and Matthews and Yip (1994) for Cantonese. Where page numbers are given without citation, these are the sources of the data.

### 3.1. *Final Consonants*

Mandarin retains only two consonants, *n* and *ŋ*, as possible finals in a syllable. Cantonese retains six (the full complement from Middle Chinese)—voiceless stops and nasals at three points of articulation each. This renders Cantonese's phonotactics more complex in this respect than Mandarin's.

### 3.2. *Tones*

Mandarin has four lexical tones; some dialects have only three, such as Línchéng in Héběi (Norman 1988, 194).

Cantonese has six tones by conservative estimate, and nine if one includes the three tones "checked" by unreleased consonants. Moreover, in Mandarin, unstressed syllables lapse to neutral tone; in Cantonese, tone is assigned obligatorily regardless of stress. Thus in Mandarin, affixes and particles lack tone, while in Cantonese, each of such items' tone must be acquired on a case-by-case basis, rendering it less predictable than Mandarin in this respect: for example, Cantonese's experiential aspect marker *gwo* versus progressive aspect marker *gán* or adverbial *dāk* (22).

### 3.3. *Tone Change*

Mandarin has three sandhi rules: 1) third-tone changes to a low tone when followed by a nonthird tone, and 2) to a rising tone when followed by another third tone; meanwhile, 3) rising tone changes to a high-level tone when preceded by either a rising or high-level tone and is followed by any of the four tones (as opposed to a neutral one) (8). There are a few cases, most kinship terms, where the second rule fails to apply (e.g., *bǎobǎo* [*\*báobǎo*] "little precious," *jǐějiě* "elder sister" [Norman 1988, 150]).

Cantonese tone change is conditioned by morphology and semantics, in a wider range of contexts and with less regularity than Mandarin's more mechanical phonologically conditioned troika of rules. In Cantonese reduplications and compounds, the tone of a second element with a low-register tone becomes a mid/high rising tone (23).

This process is regular in reduplications with an AAB pattern where B is a modifying suffix:

- (2) fèih-fèih-déi > fèih-féi-déi  
fat-fat-APPROX  
'rather fat, chubby'

It is less regular in lexicalized compounds, in instances like the following:

- (3) jó yauh → jó-yáu  
left right  
'about, approximately' (23)
- (4) jouh yaht → jouh-yát  
work day  
'work day shifts' (26)

But then while there is no change in *góng daaih wah* "tell lies," the change occurs in *góng siu wá* "tell jokes" (27).

The tone change in compounds can be seen as indexed to a general denotation of conventionalization, and in a further symptom of this, rising tone change also encodes familiarity in reduplicated constructions—but in a fashion hewing to certain typical patterns although unpredictable (25). For example, *daih-daih* > *dàih-dái* "younger brother," but *taai-tai* > *taai-tái* "wife." In some cases, the tone change is other than the rising one: *bā-bā* > *bàh-bā* "father" versus "daddy" (this contrasts with the analogous Mandarin cases in which the exceptionality consists solely of the failure of the relevant rule to apply, rather than developments idiosyncratic to the lexical item). This familiarity-conditioned tone change also occurs, albeit unpredictably, in compounds, such as *hauh mùhn* "back door" > *hauh mún*, which I take the liberty of rendering as "that ol' back door."

Meanwhile, where reduplication denotes diminutiveness or tentativeness, the tone of the first rather than second element rises (due to the contraction of intermediating *yāt* "one"):

- (5) si yāt si → sí-si  
try one try  
'have a try' (23–24)

The rules for tone change in Mandarin are almost exclusively regular ones located in the relatively predictable realm of phonology. They are thus all but barred from the capricious irregularities of the semantic and pragmatic modules, which inherently render Cantonese's tone change less predictable. This quality is only enhanced by the complications introduced by the elision of *yāt* as well as the irregularities in tone change in terms of address.<sup>3</sup>

### 3.4. *Affixes*

The definition of affix is thorny in Chinese. Many of the world's languages have batteries of derivational morphemes sharply delineable in being tightly bound to stems

and phonetically unrelatable to any other roots in the language (e.g., English *-ness*, *-hood*). In Chinese, these derivational meanings are carried by a continuum of morphemes, ranging from lexical items used often with derivational meaning in compound constructions to a small number of “real” affixes, with no lexical source in the modern language, classifying more intuitively as affixal in the cross-linguistically traditional sense.

My counts here are based on an attempt to address the derivational items which, despite the clinal nature of their lexicality, most linguists would likely see as leaning toward the grammatical realm upon that cline. This excludes items that key grammars’ authors classify without question as forming compounds rather than words, such as those Chao (1968) designates as “versatile end morphemes.” However, it includes items that the same authors classify as more affixal even when they have lexical reflexes in the grammar, such as those conventionally treated as translations of European suffixes like *-ology*. Cases such as these tend to exhibit a degree of productivity, including occurring with a wide and flexible range of relevant constituents, which are generally considered significant in defining affixation from compounding.

There is inevitably a degree of arbitrariness in where the line is drawn here, given the continuum nature of the manifestation of compounding and derivation in these languages. However, since all natural languages have derivation, few persons would disagree that some segment of the continuum indeed comprises derivation rather than compounding, even if there are ample gray zones. I believe that within this segment, although opinions will differ as to exactly how much of the continuum it comprises, cross-varietal Sinitic comparison reveals an important qualitative difference.

According to this metric, Mandarin has six derivational prefixes (37–38) while Cantonese has eleven (32–34).

Mandarin has about sixteen suffixes by my count. Two are no longer productive, *-er*, and *-zi*, and *-er* is a highly unstable morpheme largely preserved via prescriptive forces, rare to absent in many Mandarin dialects outside of Beijing (Barnes 1977). Another suffix, *-tou*, is largely unproductive, except in one particular meaning (*-tou* means “worth” in constructions like *kàn-tou* “worth seeing” [Li and Thompson 1981, 43–44]). Li and Thompson analyze these latter three as a compensation for the homophonies created by the tonal levelings and final consonant erosions in Mandarin, rather than as fossilized remnants of once-transparent derivations. Meanwhile, Chao (1968, 244) gives *-ba* “attached, appended,” and twelve of what he terms “modern” suffixes, many of which encode functions carried out by Latinate affixes in English (e.g., *-ize*, *-ity*, *-ist*) as well as meanings such as method, member of group, and rate of occurrence (ibid., 225–45). (To be as representative of Mandarin as possible, I have referred to Chao’s list despite the much smaller one in Li and Thompson and other Mandarin descriptions.)

Cantonese has, by conservative count, twenty-one suffixes. These carve up the semantic field more finely than Mandarin’s, such as gender-distinct items *-láu* and *-pòh* for colloquial designations of individuals (*yáuh-chín-láu* have-money-guy “rich guy,” *baat-pòh* eight-woman “nosy lady, gossip” [37]), and Cantonese’s suffix for particular specialization (*yúhyih-hohk-gā* “linguist” [35]) as opposed to its more general “practitioner” suffix (*gaau-sī* “teacher”). Moreover, Cantonese’s cognates to

Mandarin's nonproductive *-zi* and *-tou* occur less than their Mandarin equivalents. Thus somewhat less of Cantonese's affixation component hovers below processibility as affixal for the speaker, rendering the system that much more robust.

### 3.5. *Classifiers*

Mandarin uses noun classifiers with numbers, demonstratives, and variably with quantifiers (104–5):

- (6) (a) sān ge rén  
three CL person  
'three people'
- (b) zhèi zhǎn dēng  
this CL lamp  
'this lamp'
- (c) zhěng ge fángzi  
whole CL house  
'the whole house'

In spoken Mandarin, the general classifier *ge* tends to supplant many of the more specialized ones, these largely tokens of formal writing (112). Yu (1945 [cited in Hashimoto 1986, 93]) found that nineteen of twenty native Beijing speakers used *ge* for "one tail" (*yī ge yī-bā*) rather than the "proper" *tiao* (slender/flat) or *ger* (slender but not necessarily flat).

Cantonese uses classifiers not only in the same grammatical contexts as Mandarin, but also extends their usage to give overt expression to referentiality, while Mandarin leaves this distinction to context, hewing closer to the telegraphic tradition in Chinese as a whole. Here is an example of classifier-encoded referentiality in Cantonese (93):<sup>4</sup>

- (7) (a) Yuhng hùhng bāt sé hóu dī.  
use red pen write good a-bit  
'It's it better to write with a red pen.'
- (b) Jī bāt hóu hóu sé.  
CL pen good good write  
'That pen is good to write with.'

### 3.6. *Aspect Markers*

Mandarin has five aspect markers according to traditional documentation (185–236):

perfective *le*  
experiential *guo*  
progressive *zài*  
continuous *zhe*

delimitative encoded with reduplication, with the reduplicated elements optionally mediated by *yī*

Cantonese has six aspect markers (197–210):

perfective *jó*  
 experiential *gwo*  
 progressive *gán*  
 continuous *jyuh*  
 delimitative *háh*  
 habitual *hōi*

This count, it should be noted, is a conservative one, designed to allow Mandarin maximal benefit of the doubt. I refrain from including the less central alternates to the core aspect markers in Cantonese that Matthews and Yip describe that do not appear to affect the overspecification quotient, or the “whenever” — marking subsidiary function of habitual marker *chān* that they call attention to as an “aspect marker” but which could also submit to adverbial classification (228).

Nevertheless, even with this conservative count, Cantonese comes out ahead in overspecification. For example, Cantonese is one of the many languages in the world that overtly expresses habituality while Mandarin does not. And furthermore, in its aspect-marking paradigm, Cantonese has transformed further along the cline of overspecification than mere enumeration of elements can suggest. Mandarin’s progressive marker *zài* is an extension of a locative verb still robust in the grammar, while Cantonese’s *gán* is a distinct item from its Mandarin “to be at” *zài*-cognate *hái*. Cantonese’s delimitative *háh* in many instances is a simple equivalent (i.e., not “more complex”) of Mandarin’s encoding the same concept with reduplication and optional *yī*, but is also used with reduplication to connote prolongation of an action, here departing from the domain of Mandarin’s construction and constituting an overt marking of a distinction that Mandarin lacks:

- (8) Ngóh lám-lám-háh,      dōu haih mhóu bŭn ũk.  
 I      think-think-DELIM also is not.good move house  
 ‘I’ve been thinking, it’s best not to move.’ (208)

### 3.7. Negator Allomorphs

Mandarin has three negators assigned according to semantic context (415–17): *bu* is the general predicative negator, *bié* is used in negative imperatives, and *méi*, optionally suffixed with *yǒu*, encodes the negative existential.

Cantonese has seven negators. The items *mh*, *mhóu*, and *móuh* are equivalents to the above mentioned Mandarin forms, respectively. But there are also, for one, *meih* for “not yet” and *míj* for “not only.” Then, with *dōu* “all,” *mhaih* is used:

- (9) Daaihhoik mhaih bīngō dōu sāu ga.  
 university NEG-be who all accept PRT  
 ‘The university doesn’t accept just anyone.’ (262)

There is an additional negative imperative marker *máih* that connotes familiarity with the addressee:



- (10) (a) Mhóu sihk lā, gam làahnsihk.  
 don't eat PRT so bad.eat  
 'Don't eat it, it tastes so bad.'

but

- (b) Máih sihk lā!  
 NEG eat PRT  
 'Don't eat that!' (364)

Cantonese requires overt attendance to more shades of the negative than Mandarin.

### 3.8. *Complementizers*

Mandarin uses the item *de* in three functions united semantically by a core concept of adverbial modification:

- (11) (a) manner:  
 tā chàng **de** hǎo tīng.  
 he sings DE nice hear  
 'He sings beautifully.' (85)
- (b) extent:  
 chǎo **de** rénjia shuì bù zháo  
 make.noise DE other.people sleep NEG achieve  
 'make so much noise that others cannot sleep' (86)
- (c) potential:  
 kàn **de** wán  
 read DE finish  
 'can finish reading' (86)

But Cantonese encodes manner with a different morpheme than it does extent and the potential:

- (12) (a) manner:  
 Kéiuh hohk **dāk** hóu faai.  
 he learn ADV very fast  
 'He learns fast.' (179)
- (b) extent:  
 Kéiuh fan **dou** m̀h jī séng.  
 she sleep till NEG know wake  
 'She slept so well that she forgot to wake up.' (156)
- (c) potential:  
 Géi dím fēi **dou** Sāam Fàahn Sīh a?  
 how time fly arrive San Francisco PRT  
 'What time do we arrive in San Francisco?' (218)

Here, Cantonese lies at an intermediate point between two extremes: Mandarin with *de* in all three functions and Southern Min with not three but four items

divided among those functions, subdividing the extent domain between two items according to a fine semantic distinction (Lamarre 2001, 113–14). Nevertheless, between Mandarin and Cantonese, Mandarin is the less overspecified variety.

### 3.9. Sentence-Final Particles

If sentence-final particles are defined in Mandarin as taking neutral tone and not being reflexes of homophonous items used elsewhere in the grammar, Mandarin has six conventionalized ones (238). Chao (1968) is more exhaustively descriptive than Li and Thompson, who are as much concerned with addressing structuralist debates as including every jot and tittle in the language. Yet despite Chao's looser characterization of particle, applying the above definition to his more copious list (*ibid.*, 801–9) yields the same number.

Cantonese has about thirty sentence-final particles, with tones assigned to them specifically (338–39). Furthermore, their combination is not free, and many occur in specific places in a sequence, for example, only first or only last. As a result, Cantonese gives conventionalized segmental (as opposed to intonational or contextual) expression to a vaster array of pragmatic meanings than Mandarin, with the added structural elaboration of the constraints on combinability and ordering among the particles. A useful comparison is the two sentences below, both apprising the addressee of a fact they have not acknowledged. Cantonese has a string of four particles where Mandarin needs but one.

- (13) (a) Cantonese  
 Kéuih lô-jó daih yāt m̀hng **t̀im ge la wo**.  
 she take-PERF numberone place too PRT PRT PRT  
 'And she got first place too, you know.' (345)  
 (*t̀im* is evaluative, *ge* is assertive, *la* denotes currency, *wo* newsworthiness)
- (b) Mandarin  
 Wǒ hái děi xiě yī piān lùnwén **ne**.  
 I still must write one CL dissertation PRT  
 'I still have to write a dissertation.' (in response to a person's remark that the speaker doesn't have anything to worry about anymore) (303)  
 (*ne* translates roughly as 'Mind you')

## 4. Second Demonstration Case: Min

The facts are similar to varying degrees in comparisons between Mandarin and each of the other Chinese varieties. Studies showing that the nonstandard languages are more complex than Mandarin in one or more features constitute a healthy strain in the Siniticist literature, while studies showing Mandarin as more complex (i.e., Li 2001 on Mandarin's preposition inventory) constitute a mere trickle.

Min is useful for a briefer supplementary comparison to the Cantonese one, given that it is relatively well covered in the literature and is overall even more complex compared to Mandarin than Cantonese.

4.1. *Phonemic Inventory*

Min dialects have on average eight vocalic contrasts to Mandarin’s usual five, with the Yongan dialect having as many as fourteen (ten oral, four nasal) (Norman 1988, 237). Many Min dialects preserve all six original final consonants (*ibid.*).

4.2. *Phonology of Proto-Min*

Norman (1973) argued that the nature of the tones in Northwestern, Northeastern, and Far Western Min require that Min developed from a proto-language with a five-way manner distinction in initials rather than the three-way one traditionally reconstructed for the Middle Chinese that was ancestral to Mandarin. Crucially, he supposes that this five-way distinction was likely the result of erstwhile consonant clusters or prefixes that would be ungainly to reconstruct as developments from the Middle Chinese phonology or morphology.

Thus Norman’s argument suggests that in its development from Old Chinese, Proto-Min broke off before Middle Chinese’s slight abbreviation of Old Chinese’s phonetic inventory, going on to develop a more complex onset system than Mandarin’s ancestor ever did. This leaves the question as to why Mandarin did not complexify phonologically in such a fashion along its transition from Old through Middle Chinese and on to the modern language.

4.3. *Tones*

Min dialects typically have seven or eight tones to Mandarin’s four (Norman 1988, 237).

4.4. *Tone Change*

Sandhi is much more elaborate in many Min varieties than in Mandarin or even Cantonese (Norman 1988, 239). Most syllables have a different tone when followed by another syllable than in isolated citation. (This is also true of Wu varieties.) Thus in Chaozhou Min (Li and Thompson 1990, 815–16):

Table 5.1. Sandhi in Chauzhou Min

	one [tsek]	CL [tuŋ]	meal [puŋ]	“one meal” <sup>5</sup>
isolation	5	3	11	
combination	3	53	11	

This process, moreover, is sensitive to syntax as well as phonology. In, for example, the Amoy or Hokkien variety of Southern Min, sandhi largely occurs within constituents (XPs) but is blocked between them (= sandhi proceeds; # sandhi blocked):

- (14) (a) yi [sia = k'a kin]<sub>VP</sub>  
 he write more fast  
 'He writes faster.'
- (b) [yi sia]<sub>S</sub> # k'a kin  
 he write more fast  
 'It would be faster for him to write.'
- (M. Chen 1987, 119)

This sensitivity to syntax is fine grained enough that sandhi is blocked between verbs and their arguments but not between verbs and their adjuncts (Chen 1987).

#### 4.5. Negator Allomorphs

Southern Min has nine negator allomorphs to Mandarin's three, subdividing this semantic space much more finely (Chappell 2001a, 346):

<i>bô</i>	perfective
<i>m̄</i>	imperfective
(iá) <i>bē</i>	negation of expectation
<i>boē</i>	negation of ability
<i>boài</i>	negation of perfective desiderative ("didn't want to V")
<i>m̄mài</i>	negation of imperfective desiderative ("don't want to V")
<i>mài</i>	negative imperative
<i>m̄mó</i>	negative hortative
<i>m̄bién</i>	negation of necessity

Cheng (1977), meanwhile, shows that when these morphemes are used as sentence-final particles, they render the semantics of questions more finely expressed overtly, the particles linked to distinctions such as assertive/nonassertive and shades of modality, in contrast to Mandarin's single *ma*:

- (15) I ē-lâi            **bó/bé?**  
 he likely-come NEG  
 'Is he likely to come?'
- (16) I ē-lâi            **m̄?**  
 he likely-come NEG  
 'He is likely to come, isn't it so?' (168)

#### 4.6. Complementizers

While Mandarin covers the manner, extent, and potential domains with one complementizer marker, and Cantonese subdivides the domains between two, Southern Min dialects subdivide the domain between four markers, using [tio-ʔ] for manner connected with latent ability and [liáu] for manner when result is emphasized (and *.ka* for extent; *e-tit* for potential) (Lamarre 2001, 113–14).

4.7. *Semantic Overspecification in Verb Strings*

Li (1986) describes Southern Min as requiring more elements in verbal strings than Mandarin, giving overt expression to shadings of a proposition that Mandarin leaves to context:

- (17) (a) Mandarin  
       bù   xiǎng (yào) mǎi  
       NEG want will buy  
       ‘will not buy’  
       (b) Southern Min  
       bô       sĩũ   b(o)eh(ài) bé (\*bô sĩũ bé)  
       have-not think will love buy  
       ‘will not buy’ (397)

Li also observes (400) that where *yào* is deletable in Mandarin if a verb follows, *b(o)eh* in Southern Min is not.

4.8. *Resultativity Marking*

Through placement of aspect markers and negators with verb concatenations, Southern Min can “undo” the unitary semantic connotation of the concatenation to highlight the semantic contribution of one member (Li 1986, 403–4). The result is a finer-grained overt marking of the resultative.

For example, both varieties have a concatenation of the verbs *wear* and *tear* which without aspect marking signifies “to wear out”: Mandarin *chuān pò*, Southern Min *tshiēng phò*. But while Mandarin can only postpose aspect marker *le*, Southern Min can place its cognate *ũ* either between the verbs or before both, in the process expressing two shades of meaning. When interposed, the perfective marker reanimates the literal contribution of *tear* and indicates having torn as well as worn out something; when the marker is preposed, the unitary semantics of the concatenation are preserved and the meaning is ambiguous as to whether something was torn during the wearing. Thus:

- (18) (a)
- |                           |                    |                      |
|---------------------------|--------------------|----------------------|
|                           | Mandarin           | Southern Min         |
| to have worn out and torn | <i>chuān pò le</i> | <i>tshiēng ũ phò</i> |
| to have worn out          | <i>chuān pò le</i> | <i>ũ tshiēng phò</i> |
| (and maybe torn)          |                    |                      |

Similarly, with the neutral negative marker, Mandarin’s stays in place while Southern Min’s placement is flexible and gives semantic juice to the second verb:

- (b)
- |                                   |                    |                          |
|-----------------------------------|--------------------|--------------------------|
|                                   | Mandarin           | Southern Min             |
| can/will not wear out             | <i>chuān bu pò</i> | <i>tshiēng b(o)ẽ phò</i> |
| (has been worn but can’t be torn) |                    |                          |
| can/will not wear out             | <i>chuān bu pò</i> | <i>b(o)ẽ tshiēng phò</i> |
| (and maybe has been torn)         |                    |                          |

To be sure, Mandarin “wins” over Min occasionally. Besides the patient/disposal marker, Mandarin, as argued by Ying-Che Li (2001), retains seven Old Chinese prepositions while Min retains only three; and while Mandarin has ninety-three prepositions, Min has only about half as many, with quite a few of even these being borrowings from Mandarin rather than being indigenous.<sup>6</sup> But then this author is unaware of an array of features in which Mandarin exceeds Min in overspecification *remotely numerous enough to bring the dialects to par*. The general nature of Min dialects is such that if Mandarin is Dutch and Cantonese is German, Min is Icelandic—with number of tones being but the tip of the, well, iceberg.

## 5. Other Comparisons

### 5.1. Grammatical Gender

While grammatical (as opposed to natural) gender marking is obviously not a living feature in Chinese, there are marginal manifestations of it in varieties other than Mandarin. There are twenty-two inanimate nouns that take gender markers in Hakka, twenty-eight in the Jianyong dialect of Xiang, and a handful in both Cantonese and Gan (Anyi dialect) (Lau 1999, 126–28). Mandarin lacks such markers, thereby exhibiting less overspecification in this area.

### 5.2. Possessive Pronouns and Alienability

Hakka has portmanteau genitive markers ( $\eta ai^{11}$  “I,”  $\eta a^{44}$  “my”;  $\eta^{11}$  “you” (sg.),  $\eta ia^{44}$  “your”) (Chappell 2001a, 351). As distinct markers of possession phonetically unpredictable from the combination of pronouns with possessive morpheme, these are verspecifications compared to the analytic strategy typical in Chinese (Mandarin *nǐ de muqin* “your mother”). Furthermore, these Hakka markers are usually used with kinship terms, and as such are a kind of inalienable possessive marking—again, a distinction qualifying as an overspecification compared to the other varieties’ possessive pronominal constructions leaving this semantic shade to context.

### 5.3. Aspect Marking

Wu (1999) reports twelve aspectual markers in Xiang. According to Zhou (1998), Xiang has “a larger number of aspect/tense markers and a wider range of grammatically encoded aspect meanings” than Mandarin, and “tends to grammaticalize more areas of aspect/tense meaning than Mandarin does.” In contrast to Cantonese’s minor numerical superiority over Mandarin here, Zhou gives eleven markers for Xiang:

<i>da</i>	past
<i>ga</i>	perfective
<i>gada</i>	perfect
<i>kelai</i>	experiential
<i>zai(goli)</i>	progressive
<i>da</i>	durative stative

<i>ji</i>	durative as continuative background
<i>can</i>	durative as “vivid” continuative background
reduplication	brief delimitative
<i>can</i>	delimitative compared to another event
( <i>da</i> ) <i>zhe</i>	trial or “for just a little while” delimitative

According to Zhou, this paradigm encodes subtle differences such as that between simultaneous events whose conjunction is unremarkable versus those whose conjunction is pragmatically unexpected:<sup>7</sup>

- (19) (a) Zhan<sub>1</sub> San<sub>1</sub> da<sub>2</sub> **ji** kou<sub>3</sub>sau<sub>4</sub> zou<sub>3</sub>lou<sub>4</sub>  
 John make ASP whistle walk  
 ‘John walked while whistling.’ (12)
- (b) Zhan<sub>1</sub> San<sub>1</sub> kan<sub>5</sub> **can** (kan<sub>5</sub> can) ku<sub>2</sub> gada.  
 John watch ASP watch ASP cry ASP  
 ‘John cried while watching.’ (13)

Generally, southern Chinese have more aspect markers than Mandarin, with central varieties intermediate in this respect (Yue 2003, 90). Also, southern Chinese overtly distinguish past tense from perfective aspect, as in the contrast between standard Mandarin and Cantonese (ibid., 90–91):

Mandarin

- (20) Nǐ chī le?  
 you eat PERF  
 Did you eat?/Have you eaten yet?

Cantonese

- (21) (a) Léih yáuh-móuh sihk faahn a?  
 you have-not.have eat rice INT  
 ‘did you eat?’
- (b) Léih sihk jó?  
 you eat PERF  
 ‘Have you eaten?’

#### 5.4. *Classifiers*

Yue (2003, 85) notes that southern Chinese have more classifiers than Mandarin does, with northern Mandarin varieties such as Dunganese, Lanzhou, and many of Shanxi and Shandong making do with just one general classifier (this also being true of Shanghai Wu). As such, Mandarin is less complex than its sisters in terms of overspecification of noun classes.

#### 5.5. *Plural Pronouns*

In most Chinese, plurality is indicated in pronouns via the cliticization of a plural clitic onto the singular pronouns, such as Mandarin’s *men* or Cantonese’s *deih*. But

in many Wu dialects, plurality is indicated differently in each person. For example, in Shanghai Wu, one suffix is used for the first person, another for the third person, and the second-person plural pronoun [na] is a contraction of the singular pronoun and a plural marker (Yue 2003, 86). In encoding plurality differently in pronouns according to person, Wu is more elaborated than Mandarin in this respect. In the Pingjiang Changshou dialect of Xiang, the third-person pronoun differs according to visibility of the referent (*ibid.*, 86).

### 5.6. Three-way Gradation in Demonstratives

Some Wu and Hakka dialects have a distinction in demonstratives between proximal, medial, and distal. The Shengxian Changle Wu dialect takes this even further, with distinct paradigms for stative versus active (used to encode movement toward the referent) (Yue 2003, 89):

TABLE 5.2. Demonstratives in Shengxian Changle Wu

	stative	active
proximal	[ku]	[kua]
medial	[løy]	[lia]
distal	[moŋ]	[moŋ fia]

### 5.7. Zero Copula

In Mandarin, it is grammatical for the copula *shì* to be omitted optionally (Yue 2003, 96):

- (22) Tā Běijīng rén.  
       he Peking person  
       ‘He is Pekingese.’

Yue notes that this is less common in Cantonese and Hakka. Thus Mandarin is less inclined to overspecify the relationship between subject and predicate.

### 5.8. Complementizers

Lamarre (2001) shows that in a cross-varietal comparison of complementizer morphemes, Mandarin, as usual, is the least specified member in terms of the division of labor between manner, extent, and potential. In the especially Mandarinized varieties—the Central ones and Hakka—sometimes a Mandarin-style uniform strategy is grammatical, but with optional alternate markers. But out of all of these languages, only Mandarin makes no overt differentiation of any kind in its marking of these three semantic domains. From Lamarre (2001, 115):



TABLE 5.3. Complementizer Morphemes  
across Chinese

	manner	extent	potential
Mandarin	<i>de</i>	<i>de</i>	<i>de</i>
Wu	<i>de/lai/delai</i>	<i>de/lai/delai</i>	<i>de</i>
Xiang/Gan	<i>qi</i>	<i>qi</i>	<i>de</i>
Xiang	<i>de/qi</i>	<i>de/qi</i>	<i>de</i>
Hakka (Dabu)	<i>de/qu</i>	<i>de/qu</i>	<i>de</i>
Cantonese	<i>de</i>	<i>dao</i>	<i>de</i>
S. Min	<i>liao/zhe/qu</i>	<i>.ka</i>	<i>e-tit</i>

6. A Bird’s Eye View

The figure below compares Mandarin with its sisters in nine aspects of grammar, all commonly discussed in the Siniticist literature.

The uniformity of white in the Mandarin column is *not* intended to indicate that Mandarin entirely lacks the features in question. Rather, the figure takes the degree to which a feature is expressed in Mandarin as a baseline, with the import of the results being how prevalently the other languages surpass Mandarin in overspecification and structural elaboration. Black indicates an especially robust presence of the feature, and gray a moderate one: for example, Min and Cantonese have a comparatively large number of tones, while Hakka, Gan, and Xiang tend to have about six. It is crucial that in none of these cases does a nonstandard variety exhibit a feature to a *lesser* degree than Mandarin.

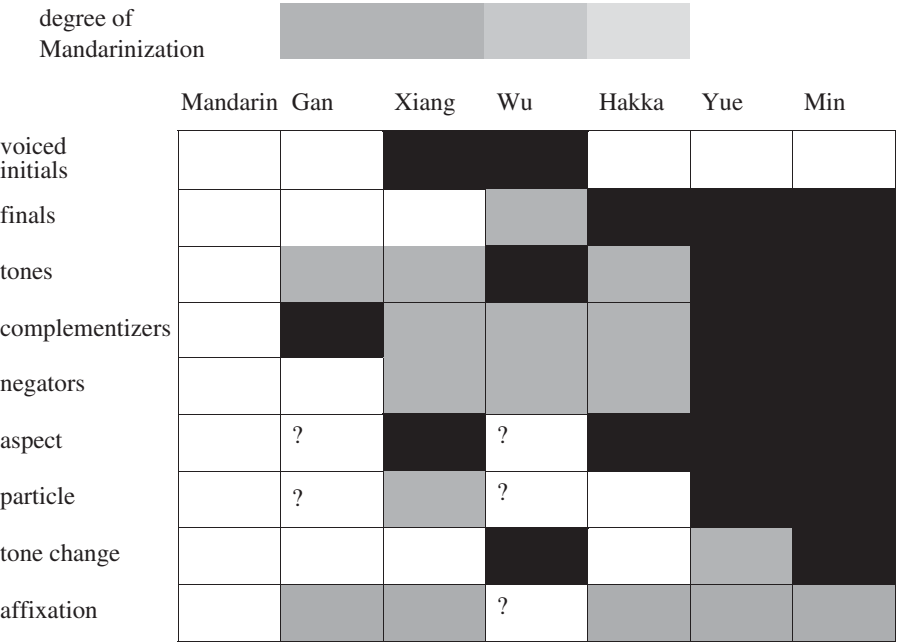
Question marks indicate that the relevant data was unavailable to me. I have made all efforts to exhaust available sources, including communications with Siniticist specialists who were gracious enough to share their knowledge (especial thanks to Hilary Chappell and Stephen Matthews). However, interest in dialectology has traditionally been rather low among Chinese linguists, such that much information has simply not been gathered to date. Moreover, there is a certain “tone fetish” in Sinitic study, such that a disproportionate amount of the literature on the lesser-studied varieties addresses tone rather than other features. Finally, some of the relevant literature is only available in Chinese, which I do not read. Engaging in comparative Chinese language research from outside of the subfield is like finding that there are full grammars of only French and Spanish, briefer ones of Italian and Portuguese, and only scattered articles on Romanian, Occitan, Catalan, and Rhaeto-Romance mostly analyzing inflectional paradigms, with the further obstacles that most of the languages’ dialects are as divergent as Italian’s and a goodly portion of the literature is written in Arabic.

These handicaps acknowledged, I believe that the figure below nevertheless captures a genuine qualitative difference between Mandarin and its sisters. Note, for example, that even in the extreme case that with further descriptions it turned out that none of the boxes with question marks could be shaded gray or black, then a striking difference in overspecification would still be evident.

Further data, it would seem, could at most strengthen but at the least only leave intact my argument.

Specifically, the figure shows that overspecification and structural elaboration increase in Chinese as one moves southeast. The distribution even captures the intermediate degree of complexification in the more Mandarinized varieties, in that gray boxes predominate in the middle columns.

FIGURE 5.1. Overspecification in Mandarin’s Sisters



6.1. Features

*Voiced initials* refers to number of voiced initial consonants retained from Old Chinese, and *finals* to number of Old Chinese final consonants retained.

*Tones* refers to number of lexical tones.

*Complementizers* refers to number of morphemes divided among the three domains of manner, extent, and potential. Black indicates a two-way distinction or more;

gray indicates the presence of alternate morphemes despite grammaticality of one morpheme in all three contexts.

*Negators* refers to number of negator allomorphs; *aspect* to number of aspect markers; and *particles* to number of sentence-final pragmatic particles.

*Tone change* refers to complexity of tone change processes.

*Affixation* refers to a combination of 1) number of derivational prefixes and suffixes; 2) number of lexical items marked by grammatical gender markers; and also 3) remnants of the \*-r- infix, reconstructed as distributive, iterative, or intensificational by Sagart (2001), exemplified by the Fuzhou Min *peɪŋ*<sup>31</sup> “to toss, turn over,” *pe<sup>31</sup>-leɪŋ*<sup>31</sup> “to turn something over repeatedly, as when frying fish” (136). Sagart finds remnants of the infix in Hakka, Yue, and Min—but not standard Mandarin. Because the morpheme is neither productive nor even perceived as morphemic by speakers, it obviously no more qualifies as an overspecification in the living grammars than *-le* in English *giggle*, *nibble*, and *wobble* does. Nevertheless, its absence in Mandarin is valuable in constituting another example of its lesser retention of Old Chinese features.

## 6.2. Data

Data used for figure 5.1 not hitherto discussed:

*Wu* preserves four voiced initial stops and some consonantal finals; *Wu* dialects have as few as five but usually seven or eight tones; it has a “not yet” negative in addition to Mandarin’s basic three (Norman 1988, 199–204); its sandhi is highly complex (*ibid.*, 202).

*Gan* preserves all six Middle Chinese finals in its conservative dialects; *Gan* dialects have as few as five but usually six or seven tones; *Gan* lacks a “not yet” negative (Norman 1988, 204–6).

*Xiang* conservative dialects preserve many Old Chinese voiced initial consonants; *Xiang* dialects have as few as four but as many as six tones; it has four principal negators (Chappell 2005). Changsha *Xiang* has nine sentence-final particles according to Bao et al. (1999, 329–33) (with profound gratitude to Yunji Wu for translation). Tone sandhi in *Xiang* is as simple as in Mandarin (Yunji Wu, Dec. 2003, personal communication).

*Hakka* usually retains all six final consonants (three-way place distinction in stops and nasals); most dialects have six tones (Hashimoto 1973, 104), while the conservative Hailu dialect has eight (Norman 1988, 226). *Hakka*, despite its southern pedigree, has rather simple sandhi (Hashimoto 1973, 111; Luo 1984, 35–36 [endless thanks to Hilary Chappell for translation]). However, it has six grammatically central aspect markers according to Chappell and Lamarre (2004). According to Henne (1966, 107), the Sathewkok variety of *Hakka* has just five sentence-final particles; despite the study’s basis on a mere 30,000-word corpus, I tentatively mark the relevant box as white in the figure.

## 6.3. Representativeness of the Sample?

Naturally, a question arises as to whether the above figure distorts the actuality in pursuit of a guiding thesis. The question is especially urgent given that the contrast

in question is not as stark as that between English and its sisters, where a simple demonstration that English has lost so much more of the Proto-Germanic legacy than any other Germanic language so strongly supports a case for “semi-semi-pidginization.” A Mandarin “English” would have, for example, no tone, no final consonants, and a single negator.

Not only is this obviously not the case but also Mandarin has its unique developments, and one could surely compose a lengthy list of them. However, this is predictable of any natural language. My thesis is that Mandarin experienced an interruption at a certain point—but certainly after this, it took on new features on all levels of its grammar: only the death of a grammar prevents it from doing so. Neither is there reason to suppose that the Altaic impact would have eliminated all of the idiosyncratic features Mandarin had been developing beforehand; for example, the *bǎ* transitive marker appears to emerge before the tenth century (during the Tang dynasty) (Norman 1988, 131).

More to the point, however, all of the other varieties have their local quirks as well, such as Hakka’s portmanteau genitive markers and Wu and Min’s baroque sandhi rules. That is, we could compose lengthy lists of unique post-Old Chinese developments for each of the languages. However, while developing their quirks, Mandarin’s sisters have simultaneously tended to complexify the features in figure 5.1, while in many cases also retaining features from Old Chinese that Mandarin shed. Thus alone, in the logical sense a list of Mandarin’s local innovations appended to figure 5.1 would leave it less complex overall than any of its sisters, since including such a list for Mandarin would require also including the various features that the other languages have elaborated. Mandarin would still stand as an anomaly, because Mandarin alone accreted its local developments while having also lost so much complexity, while the other Chinese accreted their local developments while having evolved with much less, if any, interruption.

Thus, strictly speaking, our questions are:

1. whether Mandarin has for some reason developed many *more* local features than any of its sisters have, thereby “compensating” for not having elaborified the features in figure 5.1 and coming out at par with its sisters,

or

2. whether there are a good ten or more features I have neglected—thus unequivocally dwarfing the nine in figure 5.1 into an unrepresentative abbreviation of a larger reality—that Mandarin has elaborified along with many or all of its sisters to varying degrees, with the case often being that some of the sisters are the least elaborated while Mandarin falls in the middle or extreme range.

The answer to both questions is highly likely to be no.

As to the first scenario, given that there is no “strange attractor” phenomenon that drives grammars toward a specific degree of needless complexity, the idea that Mandarin has somehow necessarily maintained a complexity equivalent to its sisters lacks explanatory foundation, and would seem to be a random surmise only attractive as a way to shore up the truism that all languages are equally complex. In any case, in the case that Mandarin had indeed for some reason undergone an un-

usual degree of internal developments, the question would still remain as to why it alone had not elaborified precisely the nine features in figure 5.1.

The second scenario is equally unlikely. For one, the features in figure 5.1 cover a fairly wide range of Chinese grammar—phonology, tonology, VP modifiers, pragmatic marking—suggesting a general phenomenon at work. It would be difficult to conceive of a reason why Mandarin would only come out as equally elaborified as its sisters in features that grammarians have traditionally paid little attention to for centuries. We must consider that, for example, careful perusal of grammars of Swedish and Dutch easily reveals them as complex to a similar degree, while the same approach to grammars of Mandarin and Cantonese just as easily reveals a major difference on this score. If Mandarin harbors a critical mass of features as complex as or more more complex than their manifestations in the other varieties, and these features are “subtle” ones somehow only gleanable via effort and ingenuity, then this alone would only confirm Mandarin’s extraordinary status.

An analogy would be the claim that creoles will reveal their parity in complexity to Serbo-Croatian when we have more comprehensive creole grammars. Given that Serbo-Croatian reveals its complexities within the first pages of even the most cursory traveler’s handbook, while creole grammars appear anomalously streamlined even in academic grammars hundreds of pages long, then if creoles will only reveal their complexities in uniquely exhaustive descriptions, this would only support an argument that they are qualitatively unique from older languages.

Less direct yet indicative support for the implications of figure 5.1 comes from the aforementioned predominance in the Siniticist literature of studies showing Mandarin as less complex than its sisters in various respects—by scholars with local interests unconnected with an analysis such as mine. Another creolist analogy beckons; namely, the wealth of creolist articles arguing that a given creole is grammatically simpler than its substrate languages, by scholars with no intention of founding a general creole genesis theory upon simplification.

## 7. What Held Mandarin Back?

### 7.1. *Degree of Anomaly*

English is unique among Germanic languages in how much of the Proto-Germanic legacy it has lost, but the difference between Mandarin and its sisters traces much less to degree of retentions from Old Chinese. For example, Mandarin, with its three negators, has abbreviated the battery of negators that Old Chinese had: general *bù*, existential *wú*, “not yet” marker *wèi*, negated copula *fēi*, and two portmanteaus resulting from the suffixation of postposed object pronouns (Norman 1988, 97–98). But then reflexes of the last three of these no longer exist in the other varieties either, paralleling assorted Old Chinese features that have barely survived in any living variety, such as the case distinction in some pronouns (e.g., first-person singular subject *wú* and accusative *wǒ*) (however, cf. the exception of Xunhua Mandarin as the result of language contact [Yue 2003, 87], along the lines of similar varieties treated below in 9.2). Furthermore, there are features robust in modern Mandarin that were

only beginning to develop in Old Chinese, such as numeral classifiers, which are documented as far back as the second century BC and were conventionalized by the fifth century AD, having proliferated while derivational affixation was wearing away (Huang 1964; Sagart 2001, 124).

Despite cases like this, Mandarin remains unique in how relatively little it has innovated compared to the other varieties. For example, the southern varieties have innovated their vast batteries of negator allomorphs, while Mandarin has simply retained its three.

Similarly, amid the vast literature on historical tonology in Chinese, a question rarely addressed is just why Mandarin stuck with only the four tones of its direct ancestor Middle Chinese. All of the other dialects developed new ones, as Old Chinese's voiced initial consonants conditioned register differences that developed into new tones, which were then phonemicized when the voicing contrasts collapsed. The question is even more pressing given that the development of three or four tones into several more is an almost universal process in Miao-Yao, Tai, Viet-Muong, and Lolo-Burmese (Norman 1988, 9). So inexorable is the force that voicing and phonation have exerted upon monosyllabic languages in Asia that, to wit, it is not "normal" for a grammar that began with four tones and two voicing-conditioned registers to maintain just four tones a millennium and a half later.

The overall difference between Mandarin and its sisters in overspecification and structural elaboration strongly suggests that something in Mandarin's time line disrupted the regular transformation that grammars undergo over time. Short of the rather unscientific choice of attributing the linguistic anomaly to chance, sociohistory beckons as a causal explanation. Namely, the facts point to a significant degree of structural abbreviation due to extensive second-language acquisition, with standardization occurring only after this abbreviation had taken place. Standard Mandarin now stands, then, as an officialized Chinese streamlined by moderate pidginization effects due to non-native speakers transmitting their rendition of the language to succeeding generations.

## 7.2. *Foreign Rulers: Garden Path?*

Here, Hashimoto and others understandably look to the fact that, historically, northern China was long distinct from southern China in being regularly not only invaded but also governed by Altaic-speaking peoples, the latter situation first occurring under the Hsiung-Nu in 304 and recurring intermittently through the Manchu takeover of 1644. However, this analysis becomes less attractive in view of modern developments in language contact theory and the generally accepted standards of proof that have been obtained in the field over the past twenty years.

Certainly these foreign rulers appear initially germane in a search for foreign speakers' influence on Mandarin. However, these invaders came to China not as migrant populations but as imperial administrators. Their presence was largely restricted to royalty, government officials, and soldiers, and they made little effort to learn Chinese or otherwise assimilate into Chinese culture.

For example, in its historical documentation, the crucial change in Mandarin appears in the early fourteenth century, at which point it displays most of its mod-

ern phonological cast, having lost all final stops and thus the entering tone they conditioned (Norman 1988, 49); this is, in fact, the point at which a recognizable Mandarin is considered to have emerged. Hashimoto (1986), noting the Altaic influence in the Mandarin of *The Secret History of the Mongols* and the *zi-di-shu* ["Son's Book"] of the eighteenth century, implies that the Manchus were the culprits, but the four hundred-year gap between the linguistic change and the sociohistorical one is rather uncomfortable. It would seem more graceful to surmise that a more contemporary occupation was responsible, and here, the alien regimes from 907 through 1368 of the Tangut, Khitans, Juchens, and finally Genghis Khan may beckon.

But, typically of foreign rulers in China, these peoples could have had but marginal linguistic influence on the mass of Chinese speakers. The Han vastly outnumbered them except in the highest realms of the government and military, and even here, the aliens largely relied on interpreters. As Franke and Twitchett (1994, 35) note, "Millions of rural Chinese may well have never encountered a foreigner," while even in the cities, foreigners were largely soldiers. To propose that distant court rule had a significant effect on the transmission of Mandarin among millions of people is as implausible as earlier claims, now refuted by Thomason and Kaufman (1988, 263–342) and others, that similar rule by the Normans affected English morphosyntax among millions of commoners who knew neither French nor French-speaking people. In the Norman case, lexicon and derivational morphology nevertheless eventually penetrated the language through writing as the invention of print soon after the Norman reign exposed the elite rendition of the language to the public. But even this was impossible under Altaic-speaking rulers in China during the period in question, as printing and literacy were elite activities largely unknown to the illiterate masses. Moreover, the lexical and grammatical Mongolisms that De Rachewiltz (1996) notes in Chinese writings of the thirteenth and fourteenth centuries were artifices due to Chinese translation having been inserted into Mongol documents interlinearly (Franke and Twitchett 1994, 32), no more reflective of spoken language than, for example, the syntactic distortions in many European languages' renditions of the Greek New Testament.

### 7.3. *A New Explanation: Massive Migrations*

We seek instead a mechanism via which new generations would be exposed as much to abbreviated Mandarin as to its full rendition, and this would require evidence that foreigners infiltrated the general population in significant numbers. Where Chinese history does offer useful evidence here is the three centuries before the foreign regimes, during the preceding Tang dynasty starting in 618. During this period, vast numbers of foreigners were incorporated into northern China in close contact with Han people.

As early as the first century AD, the entire southern portion of the empire of the Hsiung-Nu, Altaic speakers who had long threatened from the north, was incorporated into China, with large numbers of Hsiung-Nu resettled in the northern provinces of (in modern terminology) Gansu, Shanxi, and Inner Mongolia, and large

numbers of Han resettled into these regions to form mixed settlements with the foreigners (Yü 1986, 401).

But under the Tang dynasty after 618 this kind of incorporation of foreigners was especially robust. During this period the government settled hundreds of thousands of Turks, Uighurs, Khitan, Sogdians, Ch'iang, Tangut, T'u-yü-hun, and Tibetans into northern frontier provinces. These peoples assimilated to Chinese society as farmers and herdsmen, many also enlisting in the army, often serving as commanders, especially at foreign outposts (Franke and Twitchett 1994, 9). Thus in 630, after a victory over Turks in the west, the Chinese government settled about a hundred thousand Turks along the frontier from Shanxi to Hebei, with ten thousand settling in then-capital Changan (Wechsler 1979, 223). Their sons and daughters became slaves and serfs—whose possible linguistic impact is indicated by the central role that Khoi-speaking nursemaids and servants played in transforming Dutch into the semicreole Afrikaans in South Africa. Between 690 and 694, 350,000 Turks, Tibetans, and Man (Miao-Yao-speaking peoples from the southwest) migrated as refugees from local conflicts into China (Guisso 1979, 311). Tanguts driven from Tibet by Tibetans in the 600s and 700s settled in the Hsia-chou/Yin-chou region near the Great Wall at start of the 800s (M. Dalby 1979, 679).

As a result of these resettlements, during the Tang dynasty in a vast band of northern provinces stretching from eastern Qinghai eastward across Gansu, Shaanxi, Shanxi, and Hebei, non-Hans were, crucially, a majority. They coexisted peacefully with Hans living with them side by side, with a degree of intermarriage (Franke and Twitchett 1994, 9). Importantly, the area in question comprised the provinces containing early capital Changan (Shanxi province) and later capital Peking (Hebei province), sites where Mandarin was standardized. Although a number of these people spoke Tibetan rather than Altaic languages, the Tibetan group displays the same SOV word order and inclusive/exclusive distinction in plural pronouns that Hashimoto and others note as parallels between Mandarin and Altaic.

Here, then, is a context where surely legions of adults were acquiring Mandarin as a second language with varying degrees of success, while also being numerous and rooted enough to transform the input new generations were exposed to. Clearly this scenario is a more promising explanation for Mandarin's anomaly than a tiny foreign elite sequestered in major cities communicating through interpreters.

This account also squares better with the documentary evidence than one placed later in time. Typically, changes in spoken language appear in writing long after their emergence, the time lag often comprising centuries. In this light, it would be curious that a newly abbreviated Mandarin be used in writing in the early 1300s if it had emerged only in the previous century under the Mongols. Even proposing its emergence between 900 and 1200 under other alien regimes is something of a tight fit compared to assuming that Mandarin's transformation occurred during the three centuries before this. Thus the *Zhongyuan yinyun* rhyme book of the early 1300s would reflect changes that had had as long as six hundred years—roughly the span of time between *Beowulf* and Shakespeare—to spread, conventionalize, and escape the bonds of pejorative social evaluation.



## 8. Further Evidence

### 8.1. *Distributional Anomaly*

The sheer distribution of the Chinese varieties strongly supports that there was an interruption in Mandarin's natural development. A linguistic map of China reveals a clustering of highly distinct languages in the southeast that dilutes into a vast territory of relative linguistic homogeneity northward and westward—that is, Mandarin dialects. Indeed, despite the vast territory they occupy, Mandarin varieties are strikingly homogenous compared to, especially, Min, with Fujian province being the most linguistically diverse Chinese-speaking region, with dialects often changing from county to county and even village to village (Norman 1988, 188).

Crucially, this is exactly what we would not expect according to regular language family development. It is generally agreed that linguistic diversity is greatest where a family emerged, given that speakers have occupied that area for the longest time and languages have had the longest time to diverge. Thus Austronesian, a family a thousand languages strong distributed across a vast southeast Asian and Oceanic territory, is traditionally reconstructed as emerging in Taiwan, where about two dozen indigenous languages (including several lately extinct), comprising no fewer than three of Austronesian's four subfamilies, coexist within less than thirteen thousand square miles. Bantu, a notoriously uniform group, is reconstructed as emerging in northwestern Africa, where the diversity among its sister subfamilies of Niger-Congo is so high that it continues to frustrate taxonomists.

Along these lines, based on how Chinese languages are distributed, the observer unfamiliar with Chinese history but versed in language spread theory would readily assume that Old Chinese was spoken along the southeastern coast, and only spread inland and northward later.

But of course, Chinese actually spread from the north, along the Yellow River, southward. According to canonical conceptions of language spread, we would expect that the north and southwest would by now harbor a patchwork quilt of mutually unintelligible languages, at least as diverse as the southern Chinese ones if not more so. Note, for example, that the Romance languages diverged from Latin within approximately the same period that the Chinese languages developed.

Nor does geography suffice as an explanation. For example, the mountainous terrain and absence of major rivers in the Fujian province may well account for the especial diversity within Min itself. Similarly mountainous topography explains the diversity among the Rhaeto-Romance varieties in Europe (by no means mere dialects of a single language). But nevertheless, Romance varieties as distinct as French and Spanish, or even Spanish and Portuguese, developed on terrain much less conducive to isolation—while at the same time Mandarin varieties remained so akin that speakers from Harbin in northern Manchuria can today converse with speakers from 3,200 kilometers southwest from Kunming in Yunnan province (Norman 1988, 187–88).

It would appear that the only way to reconcile accepted and intuitive conceptions of language spread with the anomalous distributional facts in China is to re-

construct a decisive interruption in Mandarin's development, exerting upon it a significant deconstruction, in the absence of which it would have mutated normally. That is, it would have proliferated via elaborification and structural elaboration into an array of new languages, as replete as the southern varieties with retentions from Old Chinese, and as distinct from one another as these are due to independent variations upon their rootstock. The homogeneity of Mandarin, then, is the result of its having "started again" in regular complexification after a radical abbreviation. Mandarin can be seen as analogous to Latin at the "Early Romance" stage, before its new offshoots had diverged enough to constitute different languages.

The distribution of Native American languages in North America is an analogous case. While it is generally accepted that humans first entered the New World by crossing from Siberia into Alaska, Native American languages are in fact least diverse in Alaska and northern Canada, comprising a relatively compact number of Eskimo-Aleut and Na-Dené languages, while diversity is densest in regions far southward such as California and much of South America. Hominoid evolutionists such as Forster et al. (1996) and Oppenheimer (2003, 307–15) have argued that this was due to an "intervention" in the regular development of languages in this region, when the last Ice Age edged out human occupants. After this, there was a new and final migration from Asia, with these people's languages today having had less time to diversify than those of the ones who had crossed over before the freeze. My claim is that the otherwise mysterious linguistic homogeneity of most of the Chinese-speaking region requires a similar appeal to external intervention.

## 8.2. *Nonstandard Mandarin Varieties*

One more piece of evidence that standard Mandarin was significantly diluted is that many nonstandard Mandarin dialects display somewhat more overspecification and structural elaboration than the standard one. Standard Mandarin, then, is an artificially reified "frozen" rendition of the language as foreigners abbreviated it, while the nonstandard varieties were freer to develop in the regular fashion. To the extent that even these nonstandard varieties are still markedly less complex than, for example, Min ones, presumably impact from standard Mandarin has put a brake on their natural development, analogously to Mandarinization restraining contiguous Chineses like Wu, Xiang, Gan, and Hakka from the elaborification Yue and Min were free to accrete. A comparison would be the more "card-carrying Germanic" cast retained by many nonstandard English dialects after the Viking impact had decisively streamlined the dialects that became standard English.

For example, many Mandarin dialects have more than four tones. Hangzhou has no fewer than seven, such that it was previously classified as a Wu dialect (Simmons 1992; Baxter 2000, 106–8). In the Jiang-Huai region five-tone dialects are not uncommon, with six-tone ones reported on the Northern/Central boundary (Norman 1988, 194). These represent a retention of one of the original four tones of Middle Chinese (the *rù* tone), as distinguished from the more common Mandarin trait of having lost this tone while collapsing the two-way register distinction between the three others into a four-tone contrast not contingent upon register (Norman 2003,

77–78). These nonstandard dialects can be seen as representing, if only partially, Mandarin’s “normal” development, arrested in the variety that was heavily impacted by second-language acquisition and then established as the standard.

In standard Mandarin, plurality is indicated in pronouns via the cliticization of *men*: *wǒ* “I,” *wǒmen* “we.” However, in many nonstandard Mandarins, the plural marker has coalesced with the pronoun, sometimes to the point of creating new morphemes phonetically unpredictable from the combination of pronoun and clitic, such as the Shanyin variety of Shanxi in which the first-person plural pronoun is [ua] and the second-person plural one is [niəu] (Yue 2003, 86). That is, these varieties have developed monomorphemic plural pronouns in contrast to standard Mandarin’s transparent derivation of plural pronouns via an invariant plural marker.

Various nonstandard Mandarins have the three-way distinction in demonstratives (proximal, medial, and distal) also found in some other Chinese (Yue 2003, 88–89). Ogawa (1981) proposes that this distinction traces back to Proto-Sino-Tibetan, and that standard Mandarin was subject to especially strong influence from Altaic languages that lacked the feature, leaving it most robust in the nonstandard dialects. The only Mandarin dialect documented to have an overt distinction between past and perfect is the Huoxian variety of Shanxi Mandarin (Yue 2003, 91).

The *\*-r-* infix mentioned above has remnants not only in Hakka, Yue, and Min but also in Shandong Mandarin and the Jin varieties of Mandarin (Sagart 2001, 141–42). This is evidence of standard Mandarin undergoing a degree of streamlining beyond what some nonstandard dialects did.

Finally, Lamarre (2001) shows that many nonstandard Mandarin varieties divide the manner, extent, and potential domains between two complementizer morphemes as more southerly varieties do, again suggesting dialects elaborifying in the “normal” way while the standard was frozen in an abbreviated state.

TABLE 5.4. Complementizer Morphemes in Nonstandard Mandarin

	manner	extent	potential
Changli (Northern)	<i>.ti</i>	<i>.ti</i>	<i>liao</i>
Xining (Northwest)	<i>zhe</i>	<i>zhe</i>	<i>lia</i>
Chengdu (Southwest)	<i>de/lai/delai</i>	<i>de/lai/delai</i>	<i>de</i>
Anqing (Jiang-huai [“Southeast”])	<i>zhe</i>	<i>zhe</i>	<i>zhe</i>

## 9. Altaicization versus Reduction

Remaining to address is the apparent consensus among Siniticists that the difference between Mandarin and its sisters is due simply to its having undergone transfer from Altaic languages rather than from the Tai, Miao-Yao, and other languages that impacted the southern varieties.

9.1. *The Altaic Legacy*

Surely the Altaic (plus, under my analysis, Tibetan) transfer occurred. Hashimoto (1986) notes that the frequency of modifier-head word order in Mandarin coexisting with head-modifier order reflects the modifier-head typology of Altaic languages (which are SOV). Thus the “native” order as in 23 (below) coexists with 24 (below).

- (23) Tā chi le mei you?  
he eat PERF NEG PERF  
‘Did he eat/He has eaten?’ (Hashimoto 1986, 88)
- (24) Nǐ bǐ tā gāo.  
you than him tall  
‘You are taller than him.’ (ibid., 85)
- (25) Tā cóng Zhōngguó lái le.  
he from China come PERF  
‘He has come from China.’ (Li and Thompson 1981, 24)
- (26) Tā (cóng) míngtiān qǐ shàngbān.  
he from tomorrow from go.to.work  
‘He will go to work from tomorrow on.’ (Dryer 2003, 48)
- (27) Tā kuài-kuài-de zǒu.  
he quickly walk  
‘He walked quickly.’ (ibid., 49)
- (28) Wǒ gěi nǐ de shū.  
I give you REL book  
‘The book that I gave you.’ (Li and Thompson 1981, 117)
- (29) Tùzi de ěrduō  
rabbit GEN ear  
‘The rabbit’s ear’ (ibid., 113)

Dryer (2003) notes that modifier-head order traces back to Proto-Sino-Tibetan, and argues that Mandarin has retained rather than innovated this order. However, he notes that Mandarin’s having retained this order to such a degree in contrast to its sisters in Sinitic and beyond must be attributed to the fact that its Altaic neighbors harbor the same order.

Hashimoto also traces the inclusive/exclusive distinction in the Mandarin first-person plural pronoun (*zámen/wǒmen*) to a like configuration in Altaic. It bears mentioning, however, that this feature is less conclusively diagnostic of contact influence than often thought, given that Norman (1988, 157–58) and Yue (2003, 87) document this feature not only in varieties of Chinese languages influenced by Mandarin such as Xiang, Gan, and Wu but also in not some, but most, Min varieties.

Along the same lines, it is generally agreed that southern Chinese varieties display transfer from languages spoken southward. Thus Bisang (1999) notes that the

use of classifiers in possessive and relative constructions in varieties like Cantonese has parallels in Tai and Miao-Yao:

- (30) (a) léih ge / go pàhngyáuh  
           you POSS CL friend  
           ‘your friend’ (MY 108)
- (b) Léih béi ngóh gó dī séung hóu leng.  
           you give me those CLpicture very nice  
           ‘The pictures you gave me are very nice.’

Ballard (1985) notes that the maintenance of voiced initials in Wu and Xiang parallels Miao and Yao, and that there are even more precise likenesses, such as that the relatively simple vowel nuclei in Wu and Xiang parallel Miao while the more complex ones in Min parallel Yao. Ballard also demonstrates that Wu’s notoriously complex sandhi operates in closely similar fashion to Miao’s, and notes work observing grammatical likenesses between Cantonese and Tai. Hashimoto (1986) notes that the ample number of tones and particles and the robustness of final consonants in the south also parallels Southeast Asian languages.

However, the accepted interpretation of these facts would appear to be the one summed up by Hashimoto (1986, 77) that “all of this leads us to the Tai or some Austroasiatic connection of the Chinese language on its southern border, and to the Altaic connection on its northern frontier.” This analysis, however, cannot account for the qualitative difference I have outlined between Mandarin and its sisters. More precisely, the data suggest that the southern Chinese varieties were impacted by extensive language shift amid proficient bilingualism, while Mandarin resulted from robust but incomplete acquisition.

### 9.2. *Altaicization to the Max*

One thing suggesting this incomplete acquisition is that the degree of specific transfer from Altaic in Mandarin is ultimately quite moderate, largely limited to a variable word order tendency and (perhaps) a pronominal distinction. Meanwhile, the parallels between the southern varieties and Tai and Miao-Yao are rich enough to lead Matisoff (2001) to group the southern Chinese within an areal “Sinosphere” shared with Southeast Asian languages.

This contrast occasions a question. If the difference in complexity between Mandarin and its sisters is due simply to the sisters having incorporated elaborified features from Tai and Miao-Yao, then why did Mandarin inherit so little of the elaborified features of Altaic? Why is Mandarin not as deeply Altaicized as the southern varieties are “Sinospheric”?

One potential explanation for this discrepancy would be to suppose that it would have been for some reason impossible, or unlikely, for Altaic to impact a grammar as deeply as Tai and Miao-Yao. Namely, we might suppose that analytic languages such as Tai and Miao-Yao offered readier sources for transfer than the agglutinative Altaic ones, given that free morphemes are more readily borrowed than bound ones.

But for one, work on language contact over the past twenty years has shown that this observation captures only a tendency. Depending on the nature of the contact situation, no module of grammar is immune to borrowing. Worldwide, languages have borrowed ample amounts of affixes from others wholesale, such as in the Amazon (Aikhenvald 2001), and just as often remodel their own affixes' functions according to those in other languages (such as the Uralic influence on Slavic; cf. Thomason and Kaufman 1988, 238–51).

And with it acknowledged that boundedness hardly bars borrowing, it must also be noted that Altaic languages hardly present as imposing a barrier to borrowing or calquing as some grammars do. Altaic languages are, for example, not fusional but agglutinative, such that the phonetic form and meaning of their affixes tend to be more readily distinguishable than those in, for example, many Indo-European or Native American languages. Altaic verbal roots are not obscured by markers of person, number, or gender. In Manchu, Mongolian, and other Altaic languages whose speakers encountered Chinese, free copular morphemes occurred sentence-finally, thus readily perceptible and presumably ripe for calquing (such that Old Mandarin could presumably have placed its copula *shì* sentence-finally). Sentence-final pragmatic particles are by no means foreign to Altaic, Manchu being rather rich with them, including modal-shaded negative ones reminiscent of those in Southern Min (Gorelova 2002). Concatenational converb constructions are a hallmark of Altaic, constituting sequences of free morphemes that would presumably be readily mimicked in an Altaicized rendition of Chinese.

We need not only speculate here. Crucially, there exist many language varieties in the northern band of China heavily resettled in the crucial window between 600 and 900 that demonstrate that it is thoroughly possible for Chinese and Altaic languages to meld much more deeply than they did during the genesis of standard Mandarin.

Tangwang is a deeply Altaicized Mandarin dialect spoken in the Gansu province, basically superimposing Mandarin lexicon upon the grammar of the Altaic language Dongxiang. This includes calques on its converbs, case-marking affixes usually modeled on Dongxiang ones but expressed with Mandarin morphemes, a causative marker, the Dongxiang aspect paradigm, and its SOV word order. This sentence demonstrates the latter four features:

- (31) (a) Tangwang:  
 nɿ<sup>31</sup> lɛ̃<sup>53</sup>-ŋ v<sup>224</sup>-xa      t͡ʃ<sup>24</sup> tɔ<sup>31</sup>-ki      ʃi<sup>24</sup>-xa<sup>31</sup>.  
 that one come-COND I-OBJ know-CAUS once
- (b) Dongxiang:  
 hə iɾə-sə matə niə mətʃiə-ʎa.  
 he come-COND I-OBJ at.once know-CAUS  
 'When he comes, let me know.' (Lee-Smith 1996a, 878)

Tangwang has even incorporated Dongxiang's logophoric possessive marker, a common feature in Altaic. Given the absence of an equivalent Mandarin morpheme, it simply borrowed the Dongxiang one:

- (32) (a) Tangwang  
 ni<sup>224</sup> lo<sup>224</sup>kuə<sup>24</sup>-xa-nə ʃuə<sup>24</sup>.  
 (b) Dongxiang  
 tʃu kakə-tə-nə kʰiəliə.  
 you elder brother-OBJ-REFPOSS speak  
 (You) talk to your (own) elder brother (about something).’ (ibid.)

Hezhou, also spoken in the Gansu province, is a similar case, with Mandarin lexicon in a grammar based on Dongxiang and the other Altaic languages Bao’an and Salar, as well as Tibetan.<sup>7</sup> Again, Hezhou, with its six case markers and other Altaic-modeled features, has sampled Altaic much more deeply than standard Mandarin:

- (33) (a) Mandarin  
 Ni yong sheme zhaodai tamen?  
 you use what serve them  
 (b) Hezhou  
 Ni tʰam-xa ʃima-la kʰuεε-tε-li?  
 you them-OBJ what-INST wait on-INTENTION  
 (c) Uighur (Turkic [subfamily of Altaic])  
 Sən ularni nimε-bilən kyt-məktʃi-sən?  
 you them-OBJ what-INST wait upon-INTENTION-2S  
 ‘What would you serve them with (i.e., food and drinks)?’ (Lee-Smith 1996b, 868–69)

A particularly striking demonstration of how deeply Chinese can be incorporated into a hybrid language is Wutun of Qinghai province, which mixes Mandarin, Tibetan, and Mongolian (Bao’an). Wutun is not only agglutinative, but is even ergative like Tibetan, and incorporates Tibetan quirks such as its evidential marking:

- (34) ɣa’mθ jita’tsɿ ji’ta’-kʰan-tœ.  
 we all of us together-look-ATTITUDE.1S  
 ‘Let us look at it together.’ (Lee-Smith and Wurm 1996, 890)  
 (normal = Mandarin, bold = Tibetan, underline = Mongolian)

There is further evidence that standard Mandarin’s degree of Altaic transfer is hardly the maximal possible, in that nonstandard Mandarin dialects often display more Altaic influence. Hashimoto (1986) gives samples of Lanzhou Mandarin with modifier-head word order foreign to the standard such as:

- (35) (a) standard Mandarin:  
 gei wò shu  
 give me book  
 ‘give me books’  
 (b) Lanzhou:  
 bă shū gěi wǒ  
 ACC book give me  
 ‘give me books’

Similarly, Dwyer (1992) shows that in the Linxia dialect of the northwest, where Mongolic languages like Bao'an and Santa have a comitative/instrumental suffix derived from a concatenation of "two" and "together," Linxia Mandarin has one derived from Mandarin's "two" plus the classifier *ge*, and also can substitute the equivalent borrowed morpheme *-la* common in many of the surrounding Altaic languages (166–69). Dwyer also describes a recruitment of Mandarin conditional morpheme *yàoshì* as a suffix *shì* used cognately to equivalent suffixes in Mongolic and Turkic languages of the area (169–71). Another example is that the inclusive/exclusive distinction in first-person plural pronouns that Hashimoto refers to is less obligatory in Beijing Mandarin than in many nonstandard dialects.<sup>8</sup>

### 9.3. *Intra-Chinese Contact: Abbreviation Minimal*

Contact, then, can add substantial new overt resources to a Chinese variety far beyond word order and isolated, highly local transfers. A related demonstration is that even when speakers of other Chinese varieties acquire Mandarin, they have tended to transfer their native structures into their rendition of the language to an extent more transformative than the light influences Hashimoto traces to Altaic, and often with no concurrent "pidginization" in any sense.

Chen (1983) documents a fifth tone in Min speakers' rendition of Mandarin in Singapore, where contact between Mandarin's four tones and Min's seven or eight results in a compromise rather than, for example, a three-tone Mandarin. Kubler's (1985) study of Taiwanese Mandarin is often described as showing a "creolization," but Kubler's terminology can only be seen as referring to the transfer aspect of creolization rather than the structural reduction aspect. This is because Taiwanese Mandarin is distinct from the Beijing variety in having a degree of transfer of Min features—dominating the page count in Kubler (*ibid.*)—far surpassing the totality of Altaic features in standard Mandarin. A related observation is that in the Xiang variety heavily influenced by Mandarin contact that Wu (1992) describes, the Xiang variety's seven tones are maintained rather than being worn down toward Mandarin's four (*ibid.*).

These cases only make it clearer that standard Mandarin could have incorporated a much vaster degree of transfer from Altaic (or Tibetan varieties), and that structural abbreviation is hardly an inevitable result when Mandarin encounters nonnative speakers. The striking degree of reduction that standard Mandarin underwent in contact with other languages, then, requires explanation.

### 9.4. *The Eternal Lightness of Mandarin Altaicization*

In that light, we must note that where Altaic *lacked* a feature (e.g., tone), standard Mandarin may have abbreviated it, and where Altaic had *an alternate manifestation* of one of Mandarin's features (i.e., head-modifier order), Mandarin may have calqued it (albeit here only partially), but where Altaic had a feature *entirely foreign* to Mandarin, standard Mandarin only rarely incorporated it (e.g., sentence-final copula, verb concatenation, or perhaps the inclusive/exclusive distinction in the first-person plural pronoun, and here only optionally). This contrasts with:



1. the truly thorough Altaicization that some nonstandard Mandarin dialects like Tangwang and Hezhou have undergone, to an extent rendering their very classification as Chinese questionable;

2. the tendency for exported Mandarin varieties to readily elaborate rather than reduce or reorganize when encountering new features in other Chinese languages with which they come into contact; and, especially;

3. the proliferation of tones, explosion of sentence-final particles, and extension of classifiers into new domains of reference that the southern Chinese underwent in contact with foreign grammars, standing in stark contrast to Altaic's rather minor impact upon standard Mandarin, rather than as a mere parallel.

Thus the Altaic parallels that Hashimoto, reinforced by Dryer (2003), has brought to attention reveal that Altaic speakers were indeed the agents of the structural reduction. However, more precisely, they left a light imprint from their specific grammars upon Mandarin, while the broader outcome of their acquisition of the language was significant abbreviation of its equipment. This abbreviation bore an imprint less specifically Altaic than simply human: in a different context, speakers of Australian, Semitic, or Indo-European languages would presumably have left similar reductions.

Importantly, this predominance of abbreviation over transfer is a diagnostic of adult acquisition, and this is precisely the sociohistorical factor that distinguishes the history of northern from southern China. While vast populations of adults were resettled in the north, contact with foreigners in the south resulted from Han migrations occasioning gradual genetic and cultural interminglings. This kind of context leads to generations of full bilingualism (which often persists today among indigenous populations in southern China), with languages exerting transfer effects upon one another without significant overall abbreviation of the structure of either. Analogies would include the Balkan Sprachbund, or Dixon's (1997) conception of "linguistic equilibrium," under which contiguous languages "stew" gradually among one another over time, adopting one another's features but with none suffering any interruption of transmission in the process.

Hence southern Chinese varieties adopted assorted features of Tai and Miao-Yao languages while remaining as awesomely elaborated overall as Yue and Min, while Tai and Miao-Yao adopted Chinese's tonality and monosyllabicity (e.g., Matisoff 2001, 317) without becoming noticeably "streamlined" structurally in comparison to other languages in the world. Cases like Tangwang, Hezhou, and Wutun show that contact between Chinese and Altaic or Tibetan languages can yield similar results, these varieties presumably having arisen where foreigners retained their native languages for longer periods, contexts predictable given that allegiance to the Chinese empire was variable among resettled peoples along the northern frontier (Franke and Twitchett 1994, 9). More typical Mandarin varieties like that of Lanzhou, with slightly more Altaic transfer than the standard, would represent intermediate cases.

## 10. Conclusion

While the membership of southern Chinese in Matisoff's "Sinosphere" is obvious, it would be much less intuitive to conceive of standard Mandarin as part of an "Altaisphere." While the parallels between Mandarin and Altaic are clearly due to contact, they are too broad and sparse to qualify as evidence of a Sprachbund. The specifically Altaic inheritance in Mandarin can only qualify as a marginal component in the qualitative difference in overspecification and structural elaboration between Mandarin and its sisters. A larger phenomenon was the main culprit, and this was general structural reduction rather than transfer.

# Persian

## *“The Most Atypical Iranian Language”*

### 1. Introduction

Anecdotally, Persian is often described as “marvelously simple.” Of course it is a complex language as all are. But the impression of simplicity is based on features genuinely peculiar compared to most other Indo-European languages. There is no marking of grammatical gender or case. The plural is not marked obligatorily (not expressed, for example, with numbers), reminiscent here of analytic languages like Chinese and Yoruba. Whereas Indo-Iranian, in general, is distinguished in having ergativity in the past tense, this is absent in Persian.

Yet things were quite different in Persian’s ancestor Old Persian, which operated along the lines of other early Indo-European languages like Latin, Ancient Greek, Sanskrit, and Old Church Slavonic. Old Persian nouns and adjectives fell into several declension classes. There were three genders, six cases, and dual marking. Typical of early Indo-European languages, nominal inflection was complicated by a rich array of morphophonemic rules conditioned by interaction between ablaut and accent. There were inflectional paradigms for present, aorist, future, injunctive (negative imperative), conjunctive (subjunctive), and optative, these expressed with endings chosen from two sets (thematic and athematic) in combination with various suffixes between these and the stem. The imperative was expressed with its own set of endings, and middle voice was expressed with inflectional paradigms as well (Kent 1953; Schmitt 1989a; Sims-Williams 1998).

Of course, even Old Persian had eliminated some of Proto-Indo-European’s elaborification, for example having collapsed the dative into the genitive and the instrumental into the ablative while its close sister Avestan (the other substantially attested Old Iranian language) retained all eight cases. Furthermore, both Old Iranian languages had already developed an anomalous Modern Persian trait also common in its sisters, the lack of animacy marking in third-person pronominals, the subject pronouns having disappeared and left their semantic field to demonstratives.

But these minor ellipses pale in comparison to a contrast between Old and Modern Persian stark enough to be regularly termed “breakdown” by analysts. Over the same period of time, Common Slavic became Czech. Why did Old Persian become so “marvelously simple”?

## 2. The Appeal to Accent

This change has commonly been passingly attributed to the tendency in Indo-European languages for stress to be retracted from word-final syllables, leaving inflectional suffixes ripe for phonetic collapse and/or elimination over time (e.g., Windfuhr 1979, 232–33; Sundermann 1989b, 110). But these observers are largely concerned with description, with little reason to consider the broader issues of explanation. Despite being understandably ready reference points for so many linguists, French and English are not as representative of Indo-European evolution as is often tacitly supposed.

Namely, a bird’s-eye view of the fate of unaccented final syllables in Indo-European reveals the melting away of suffixal paradigms as less rule than tendency, or more precisely as a mere possibility. Baltic and Slavic preserve a great deal of Proto-Indo-European nominal morphology. Pointedly, West Slavic (Polish, Czech, Slovak, Sorbian) fixed its accent on the first syllable as early as the tenth century, and yet ample nominal morphology, complete with declensional paradigms, has survived over more than a millennium since. Armenian has fixed the accent on the penult, and yet retains a rich declensional system and robust verbal inflection (only a portion of this due to innovations). Germanic includes not only members like English and Mainland Scandinavian but also ones like Icelandic and, to a lesser extent, German, in which a considerable degree of unaccented word-final inflection has survived. In Celtic, when the accent was retracted from endings, Brythonic (such as Welsh) lost nominal inflections but retained much verbal inflection; in Goidelic (such as Irish and Scots Gaelic) even some nominal declension remained.

Thus in seeking to explain what happened to Old Persian, the traditional historical linguist’s conception of *Auslautgesetze* (“laws regarding final position”) will be of, at best, preliminary use, as these are more precisely characterized as “*Auslauttendenzen*” or even “*Auslautmöglichkeiten*” (“possibilities regarding final position”).

This is especially the case given that Persian is quite unique in the Iranian family in its degree of inflectional loss. If the *Auslautgesetze* were so ineluctable as to constitute an unremarkable account of how Modern Persian arose, then it would be one of many sister languages similarly shaved down over the millennia, but this is not the case. Rather, while all of the modern Iranian languages have undergone a considerable degree of reduction of this kind (just as the modern Indo-Aryan languages have), only Persian has done so to an extent rendering it anomalous within its family. Neither Kurdish, Balochi, Pashto, Shugni, Munji, Tati, nor any of Persian’s other sisters would elicit the characterization “marvelously simple” under any analysis. This is because among them, case marking, grammatical gender, robust allomorphy, ergativity, and other elaborificational features remain, to varying degrees. In its “simplicity,” Persian stands alone.

### 3. Persian and Pashto

To illustrate the contrast, a comparison of Persian with one of its sisters, Pashto, of Afghanistan and Pakistan, will be especially useful. Pashto is especially complexified among its sisters. An acquaintance with it will help to clarify that despite the familiarity of the developments of English and French from heavily inflected ancestors, the streamlined nature of Modern Persian is by no means an inevitable mutation of Old Iranian.

In addition, Pashto is not an isolated language. It is spoken by about ten million people natively, and is a second language to many more, threatening various smaller languages. Thus this comparison will not reveal a complexity differential between a major lingua franca and the tongue of a band of obscure mountainers—although this would be indicative in itself.

Rather, we will compare two major lingua francas, both spoken over large territories in Central Asia. Interestingly, one description of Persian describes it as “to Iranian what English is to Germanic” (Windfuhr 1987, 530), while one of Pashto written independently observes that Pashto stands to Persian in “something like the same relationship as Icelandic does to English” (MacKenzie 1987, 565). In the following section, page numbers refer to the aforementioned sources for Persian and Pashto, respectively, unless otherwise indicated.

#### 3.1. *Nominal Grammatical Gender*

Persian has no grammatical gender; Pashto retains masculine and feminine marking from Old Iranian’s masculine, feminine, and neuter.

#### 3.2. *Case Marking*

Persian has no case markers. Pashto marks four cases: direct (nominative/accusative), oblique, vocative, and a second oblique, the “prepositional” case, occurring with a few prepositions. In addition, the case marking is patterned according to five declension classes, a complication alien to caseless Persian. Concord (for gender and number as well) applies between adjectives and nouns, extending even to adjectives used adverbially such as “very” (565). Pashto, then, is more overspecified and structurally elaborated in terms of case marking than Persian.

#### 3.3. *Plural Marking*

Persian is peculiar in Indo-European in not requiring plural marking. The Old Iranian-derived suffix *-ān* is restricted to adjectivally derived plurals (*borzorgān* “the elders”) and indefinite pronominals (*digarān* “the others”) except in the literary register in which it is also used for humans (531). The general plural marker is *hā*, which can connote referentiality:

- (1) (a) Ā-hā mo’allem hastand.  
           they teacher be.<sub>3PL</sub>  
           ‘They are teachers.’

- (b) Ān-hā mo'allem-hā hastand.  
 they teacher-PL be.3PL  
 'They are the teachers.'

The plural marker can also connote amplification:

- (2) (a) Mā mehmān darim.  
 we guest have.1PL  
 'We have guests.'
- (b) Mā mehmān-hā darim.  
 we guest-PL have.1PL  
 'We have lots of/all kinds of guests.' (533)

The anomaly of this within Indo-European must be clear: the reader is asked to consider a French sentence like *\*Ils sont professeur*. Moreover, plural marking is never used with numbers (cf. Spanish *\*dos profesor* 'two teachers').

In Pashto, inflectional plural marking is alive and well. While one might argue that marking plurality itself is not more complex than marking instead its referential and amplificational components, Pashto's plural marking has differentiated into an extremely complex and variant array of allomorphs. Below are the forms for just the first declension (556):

TABLE 6.1. Plural Marking in Pashto's First Declension

inanimate masculine	<i>lāsūna</i> 'hands'
animate masculine	<i>lewān</i> 'wolves'
inanimate masculine ending in -u	<i>bāñu-g-ān</i> Persian <i>ān</i> 'eyelashes'
feminine ending in consonant or -a	<i>xwle</i> 'mouths'
animate feminine ending in -o	<i>pišogāne</i> 'cats'
animate feminine ending in -e	<i>xwāxyāne</i> 'mothers-in-law'
inanimate feminine ending in -ā or -o	<i>mlāwe</i> 'waists'

There are many exceptions to cases such as these, and then there are also many suppletive plurals, such as *mor*, *maynde* 'mother, mothers,' *lur*, *luñe* 'daughter, daughters,' *zoy*, *zāmān* 'son, sons' (556–57).

Finally, while Persian suspends all plural marking with numbers, Pashto has a separate numerative marker for certain masculine nouns, used with numbers over 'one' and other words referring to amount such as *co* 'several, how many?' (557). Obviously, Pashto exceeds Persian in structural elaboration of plural marking.

### 3.4. Morphophonemics

Old Persian retained quantitative ablaut from Proto-Indo-European, resulting in complex morphophonemics throughout the grammar. The ablaut, phonetic variations upon the basic pattern  $o \sim a \sim \bar{a}$ , could, for example, occur within paradigms, as in the root for 'land, province': locative plural *dahyu-šuvā*, locative singular *dahy-šuvā*, and nominative singular *dahyāuš* (Kent 19053, 43); accompany derivational pro-

cesses (*Mārgava* “person from *Margu*” [Sims-Williams 1998, 150]); and distinguish grammatical uses of roots, such as “to carry” *bar-a-* (present), *br-* (perfect participle) (Windfuhr 1987, 543).

In Modern Persian these morphophonemic patterns are no longer remotely productive, surviving only in fossilized remnants.

But in Pashto, such patterns, while hardly as pervasive as in Old Iranian, still form a part of the grammar (555–56). An example is the second and third declensions, here exemplified by vocalic changes in the root in the second declension adjective *pox* “ripe, cooked”:

TABLE 6.2. Declension of *Pox*  
“Ripe, Cooked”

	singular	plural
nominative	<i>pox</i>	<i>pāxó</i>
oblique	<i>pāxó</i>	<i>paxó</i>
prepositional	<i>póxa</i>	<i>paxó</i>

3.5. *Pronominals*

In Persian, a set of pronominal suffixes denotes both objects (*did-am-aš* “I saw him”), possession (*ketāb-aš* “his book”), and indirect object experiencers (*garm-am ast* ‘warm to-me is’ “I feel warm”) (535). Pashto has a separate class of possessive pronominals (558): *zə, mā, jmā* “I, me, my,” thus surpassing Persian in overspecification.

3.6. *Allomorphy in Verbal Conjugation Paradigms*

In Persian, the difference in Old Iranian between thematic and athematic verbal paradigms and their application in combination with other suffixes according to tense and mood is extinct, in favor of a single set of endings applied virtually across the board. Only in the third-person singular do the endings for the present and past differ (Lazard 1992, 142, 149). (The prefix *mi-* connotes the imperfective.)

TABLE 6.3. Present- and Past-tense  
Verb Conjugations for “to Buy”  
in Persian

	present	past
1S	<i>mi-xar-am</i>	<i>xarid-am</i>
2S	<i>mi-xar-i</i>	<i>xarid-i</i>
3S	<i>mi-xar-ad</i>	<i>xarid-ø</i>
1P	<i>mi-xar-im</i>	<i>xarid-im</i>
2P	<i>mi-xar-id</i>	<i>xarid-id</i>
3P	<i>mi-xar-and</i>	<i>xarid-and</i>

In Pashto, even within the present, endings take stress when the verb is transitive but not when it is intransitive. Then in the past, the endings are phonetically different from the present ones in both numbers of the third person, where, in the

intransitive, the stress also falls upon the ending rather than the root as in the other persons. In the transitive, the element *-əl* takes stress whether it serves as an additional suffix or as a part of the ending itself (559). This is a vast mutation from the Old Iranian system, but one much more overspecified (in its overt marking of valence) and structurally elaborated (in its distinctions between paradigms and variant marking of stress) than Persian's.

TABLE 6.4. Present- and Past-tense Verb Conjugations in Pashto

	present intr.	tr.	past intr.	tr.
1S	<i>lwéǵ-əm</i>	<i>acaw-əm</i>	<i>lwéd-əm</i>	<i>ačaw-əl-əm</i>
2S	<i>lwéǵ-e</i>	<i>acaw-é</i>	<i>lwéd-e</i>	<i>ačaw-əl-e</i>
3Sm	<i>lwéǵ-i</i>	<i>acaw-í</i>	<i>lwéd(-ə)</i>	<i>ačāwə</i>
3Sf			<i>lwed-əla</i>	<i>ačaw-əla</i>
1P	<i>lwéǵ-u</i>	<i>acaw-ú</i>	<i>lwéd-u</i>	<i>ačaw-əl-u</i>
2P	<i>lwéǵ-əy</i>	<i>acaw-əy</i>	<i>lwéd-əy</i>	<i>ačaw-əl-əy</i>
3Pm	<i>lwéǵ-i</i>	<i>acaw-í</i>	<i>lwed-əl</i>	<i>ačaw-əl</i>
3Pf			<i>lwed-əle</i>	<i>ačaw-əle</i>

Also, Persian marks the singular imperative with zero and the second-person plural imperative with the same marker used for present and past (530). Pashto has a distinct marker for the singular imperative, *-a* (561).

### 3.7. Verbal Categories

Both Persian and Pashto, like all of their sisters, have vastly remodeled the Old Iranian verbal system. As analyzed by Windfuhr (1979, 1987, 535–36), Persian distinguishes between aorist, imperfective, and perfective, with the latter two occurring in constructions interpreted as present, subjunctive, and past or conditional according to context, with a perfect construction connoting the “inferential” past, based on secondhand knowledge. Below, this is demonstrated with the verb “to go” in the third-person singular, the table reorganized from Windfuhr’s version for clarity within this presentation:

TABLE 6.5. Verbal Category Marking in Persian

	imperfective	perfective
present	<i>mi-rav-ad</i>	<i>raft-e ast</i>
subjunctive	<i>be-rav-ad</i>	<i>raft-e bāš-ad</i>
past, counterfactual	<i>mi-raft</i>	<i>raft-e bud</i>
inferential past, counterfactual	<i>mi-raft-e ast</i>	<i>raft-e bud-e ast</i>
	aorist	<i>raft</i>
	inferential aorist	<i>raft-e ast</i>



The Pashto system also hinges on an imperfective/perfective distinction, perfective generally marked with the prefix *wá-*, as in *lweḡi ba* “will be falling,” and *wá-ba-lweḡi* “will fall.” But in the present, this prefix connotes the subjunctive, while in the counterfactual it marks no semantic difference, and does not occur in the conditional. The semantics of the particle *ba* are quite obscure, distinguishing future, durativity, and the conditional (561–62):

TABLE 6.6. Verbal Category Marking in Pashto

present	<i>lweḡi</i>	‘falls, is falling’	<i>wá-lweḡi</i>	‘that it fall’
future	<i>lweḡi ba</i>	‘will be falling’	<i>wá-ba-lweḡi</i>	‘will fall’
imperative	<i>lweḡa</i>	‘keep on falling!’	<i>wá-lweḡa</i>	‘fall!’
past	<i>lwedá-</i>	‘was falling’	<i>wá-lwed</i>	‘fell’
past habitual	<i>lwedá ba</i>	‘used to be falling’	<i>wá-ba-lwed</i>	‘used to fall repeatedly’
counterfactual	<i>lweḏāy*</i>	‘were to fall’	<i>wá-lweḏāy*</i>	‘were to fall’
conditional <sup>1</sup>	<i>lweḏāy ba</i>	‘would fall’		

While the Persian system has challenged analysts, it is the less overspecified one compared to Pashto’s. In the latter, the progressive and habitual are overtly marked while left to context in Persian, and the counterfactual has an exclusive marker rather than being an alternate reading of marking otherwise used also for the past. Accompanying these factors is the no longer unitary meaning of *wá-* and *ba*.

### 3.8. Ergativity

Throughout Indo-Iranian, the Old Iranian aorist and perfect categories died out, replaced by a new aorist derived from the Old Iranian perfect participle ending in *-ta*: hence Persian stem *rav-* “to go” occurs as *raft-* in the aorist. The modern languages developed a new perfect by combining the new aorist stem, marked with perfective marker *-e*, with a *be* verb: for example, Persian *raft-e ast* “has gone.”

In all but a few of the languages, the participial origin of the aorist left behind oblique case assignment to the logical subject (i.e., “the picture was drawn by me” connoting “I drew the picture”), resulting in ergativity in the past. Pashto is typical here:

- (3) Ma        kāñay ačawálay        day  
me.OBL stone throw.PASTPART is  
‘I have thrown a stone.’ (560)

The participial origin has also left behind gender agreement with the object:

- (4) Aspa-ye wáwahəla.  
mare-3S PERF-strike.PAST.3S.FEM  
‘He struck the mare.’ (565)

Persian is one of the very few Indo-Iranian languages to have lost this ergative marking completely. Only the origin of the first-person singular pronoun *man* in Old Persian’s genitive *man*—apparent only to the academic specialist—remains as a sign.

## 4. Persian in Its Marvelous Simplicity

### 4.1. *The Data*

Persian differs from Iranian languages other than Pashto in the same fashion, varying only in degree. The figure below gives a broader view of the difference between Persian and its sisters.

Indo-Iranian languages occupy not only Iran, but spread westward into Turkey, Iraq, and Syria, and eastward into Tajikistan, Uzbekistan, Afghanistan, and Pakistan (as well as a small part of China). The languages in the figure represent the key division in Indo-Iranian between western and eastern Iranian languages.

The western ones are Persian, Balochi, Kurdish, Sivandi, Farizandi, Gurani, Sangesari, Semnani, Zaza, Tati, Gilaki, and Talishi. Of these, Kurdish and Balochi are the most widely spoken. Sivandi and Farizandi represent the dialect continuum of central Iranian varieties often treated together in the literature, while Semnani and Sangesari represent the several closely related minority varieties of northwestern Iran. These northwestern and central languages are especially impacted by Persian, which may well eventually render them extinct. Gilaki and Talishi are languages classed as “Caspian,” spoken around the eponymous sea.

The eastern languages are Pashto,Ormuri, Ossetic, Yaghnobi, Ishkashmi, Munji, Bartangi, Shugni, and Wakhi. Pashto is a major lingua franca in Afghanistan and Pakistan. Ossetic, geographically isolated in the Caucasus, and Yaghnobi, now spoken by a small community in Tajikistan, constitute the northeast Iranian subgroup, and are unique as the only relatively direct descendants of attested Middle Iranian languages (Sogdian and Scytho-Sarmatian, respectively). Ishkashmi, Munji, Bartangi, Shugni, and Wakhi are small languages representing the Pamir mountain Sprachbund.

Persian is used in the Dari dialect in Afghanistan and the Tajiki dialect in Tajikistan. While there are various differences between the three main Persian varieties on all levels, none of these render any dialect especially more elaborated than another, and as such, Modern Persian will be referred to generically throughout this chapter.

The twenty-one languages in the figure are but a subset of the forty-odd members of the family (this count depending on where one draws the line between language and dialect).<sup>2</sup> I have chosen the languages on the basis of two imperatives:

1. I have concentrated on the languages for which relatively detailed grammatical descriptions are available, this not being currently the case for a great many of the smaller and/or lesser-known varieties. (Here is the occasion to offer profound gratitude to Gernot Windfuhr and P. Oktor Skjaervø for data and insights unavailable in the published literature on lesser-known varieties.)

2. I intend the figure as immediately processible to the eye within the limitations of the printed page, rather than as a vast speckled grid giving an impression more numerical than linguistic. For this reason, I have omitted a few languages whose contribution to the figure would be identical, or virtually so, to that of very close relatives included. This includes Parachi in relation to Ormuri, Roshani in relation to Shugni, Yazgulami in relation to Ishkashmi, and Yidgha in relation to Munji. That

is, the inclusion of these languages would only bolster the contrast with Persian that I intend to illuminate.

I have marked features with gray rather than black to indicate various manifestations of intermediate vitality. One manifestation is that sources show that the feature is only attested in a subset of the language's dialects. This is the case with Kurdish, Tati, and Ormuri; in the latter language, for example, the Kānīgrām dialect is more archaic and conservative while the Baraki-Barak dialect is moribund and heavily intermixed with Dari Persian and Pashto. Elsewhere, gray connotes that the feature is attested within a more specific area of the grammar than it is usually elsewhere; for example, Wakhi only has oblique marking in the plural (J. Payne 1989, 429). The indication "ni" (no information) indicates that data was absent in sources consulted (however, given that the languages in question are both closely related to Persian and heavily affected by it, all chances are that these languages do not mark the plural with numerals).

The figure highlights that the main difference is in the degree to which Persian elides categories that its sisters encode with inflectional affixation. The difference here is qualitatively lesser, then, than that between English and the rest of Germanic or Mandarin and its sisters, in which the abbreviation extended significantly even to overspecifications not marked by morphology. Persian is not, for example, especially less complex in its phonemic inventory than most of its sisters except where many have inherited segments from contact, such as a retroflex series in Pashto borrowed from nearby Indo-Aryan languages.

This demonstrates that inflection (and the distinctions it marks) is one of the most fragile aspects of grammar in situations of suboptimal transmission, such that, for example, Afrikaans lost as much inflection as English, but retains much more original Germanic derivational morphology, overspecifications, and complexity elsewhere in the grammar such as inherent reflexives and V2. However, as I outlined in chapter 3 in sections 3.2 and 3.3, inflection alone often contributes a vast degree of complexity to grammars, such that its presence alone can render a grammar more complex overall than that of an analytic sister.

#### 4.2. *Discussion of Features*

*Case markers.* Persian lacks any case marking, but other Iranian languages tend strongly to retain at least oblique marking, this sometimes divided into two classes, such as dative/accusative and general oblique.

*Grammatical gender.* No modern Iranian language retains Old Iranian's three genders, but many retain a masculine/feminine distinction, analogously to the Romance languages and many Germanic ones.

*Gender-marked third-person singular.* Persian's *u* is gender neutral, as is the demonstrative *išān* used for humans in polite language.<sup>3</sup> Many Iranian languages retain a gender distinction in third-person pronominals, in forms derived from demonstratives (cf. 4.3).

	Pe	B	K	Si	F	G	Sa	Sm	Za	Ta	Gi	Tl	Pa	Or	O	Y	I	M	Br	S	W
case markers																					
grammatical gender																					
gender-marked 3S																					
plural affix																					
plural w/ numerals				ni	ni		ni	ni													
adj. concord																					
inflected interrogatives																					
ergativity																					

FIGURE 6.1. Persian Compared to Other Iranian Languages in Inflectional Marking

*Plural affix.* Iranian is generally a family with robust inflectional plural marking, just as elsewhere in Indo-European. Persian's sisters retain the Old Iranian suffix, in some cases having developed additional plural inflectional allomorphs alongside it. Gilaki makes some use of *hā*, but only because of heavy contact with Persian (Rastorgueva et al. 1971, 61); this is also the case in Farizandi, of which the description by Rastorgueva and Moshkalo (1997) suggests that *hā* may have marginalized the original suffix to an extent similar to that found in Persian (264).

*Plural with numerals.* Persian is hardly alone among its sisters in eliding plural marking with nouns modified by cardinal numerals; however, some of its sisters refrain from this. Some languages that do elide the plural marker nevertheless have some numeral classifiers (cf. 4.3.), qualifying as a reconstitution of giving plurality a distinct signal when the modifier is a numeral.

*Adjectival concord.* Often Iranian languages' adjectives take a single concordial inflection in conjunction with nouns even if they do not retain the full early Indo-European type of paradigmatic marking as Pashto does. This feature can even persist when nouns are no longer marked for grammatical gender, as in Balochi.

*Inflected interrogative pronouns.* Persian's interrogative pronouns *ki* "who" and *ci* "what" are unmarked for case. However, in some of the languages the cognates inflect for at least two cases.

*Ergativity.* Ergativity is expressed to various extents in Iranian, ranging from Hazārūd Tati with distinct pronominal forms for past agent, past patient, present subject, and past subject to Rāxšāni Balochi where ergativity is used as a stylistic option, and not with pronouns (Windfuhr 1989, 260). Ironically, one of the few Iranian languages that lacks ergativity is one long in intimate contact with ergative Caucasian languages, Ossetic (Thordarson 1989, 457), although even here, gender marking remains in the third person on past stems. The only Iranian languages where not even this remnant of ergativity remains are those under heavy Persian influence, such as Gilaki (Rastorgueva et al. 1971), Farizandi (Rastorgueva and Moshkalo 1997), and the Caspian language Mazandarani (Gernot Windfuhr 2004, personal communication).

#### 4.3. *Complexities in Other Iranian Languages*

Figure 6.1 is designed to highlight the extent to which Iranian languages other than Persian exceed it regarding certain general features inherently amenable to binary analysis: that is, case marking or not, ergativity or not. However, the contrast is equally clear in more clinal traits. Just as this is clear in a comparison of Persian and Pashto in terms of, for example, degree of allomorphy in suffixal verbal paradigms, it is clear in similar contrasts with other Iranian languages. Typically of grammars mutating uninterrupted over millennia, these have developed overspecifications and structural elaborations replacing those lost from Old Iranian where Persian leaves them to context, and have retained Old Iranian features to a degree that Persian does not.

Balochi maintains Old Iranian consonants to a considerable degree: Avestan “tooth” *dantan-*, Balochi *dantān* (Elfenbein 1989, 353). The causative marker is an infix which also often causes a change in the preceding vowel’s length or quality: /čəɾəg/ “graze” (intransitive), /čəɾenəg/ “graze” (transitive) (Barker and Mengal 1969, 469). Balochi also retains a four-case distinction in pronouns, such as in the first person (Elfenbein 1989, 355):

TABLE 6.7. First-Person Pronouns in Balochi

	singular	plural
nominative	<i>man</i>	<i>mā</i>
genitive	<i>manī</i>	<i>me</i>
dative/accusative	<i>manā</i>	<i>mārā</i>
oblique	<i>man</i>	<i>m</i>

Most of the Tati dialects distinguish gender in verbal endings not only in the past, as a remnant of the participial origin of the stems, but also in the present (Yar-Shater 1969, 193–206), while this is unknown in gender-shy Persian. The plural suffixes inOrmuri (Kieffer 2003, 100–110) are as allomorphically variant as Pashto’s, and more morphophonemically disruptive of the stem. Imperative suffixes in Kurdish are similarly variegated and often unpredictable (Khan and Lescot 1970, 127). Ossetic has a definite marker, segmental in the Digor dialect and marked only with stress in the Iron dialect (Thordarson 1989, 469). Gurani nouns can be analyzed as falling into three declension classes (Pireiko 1997, 173–74).

The languages of the Pamir mountain Sprachbund, represented by the five rightmost languages in figure 6.1, have some numeral classifiers, such as *ux̌t bon čiray* “eight apricot trees” in Ishkashmi’s close sister Yazghulami, with a classifier for “base of stem” (J. Payne 1989, 436). This is not unknown in the western languages, Gilaki being an example (Rastorgueva 1991, 85–87).

Elsewhere in the Pamir group, in Shugni’s close sister Roshani, past-marked verbs retain a legacy of their origin as participles in maintaining a gender distinction as in Pashto, as below where *sat* is a feminine form (ibid., 438):

- (5) az-um pa Xaray sat  
I-1S to Horog go.PAST.FEM  
‘I went to Horog.’ (438)

There is also oblique marking on both arguments of the verb in the past, rather than there being marking solely on the subject as is typical of the Iranian ergative construction. This is demonstrated below, where one may note also a trait typical of the Pamir languages: a three-way proximity gradation in demonstratives for close, midrange and far (ibid., 439):

- (6) mu dum kitōb x̌eyt.  
I-OBL this.(mid).OBL book read.PAST  
‘I read this book.’

Shugni, Roshani, and Wakhi have an unusual possessive construction utilizing both locative and the oblique case (*ibid.*, 433), exemplified by this Wakhi phrase:

- (7) day mōd-ā            um            x̌uvd  
       he mother-LOC she.(far).OBL milk  
       ‘his mother’s milk’

There is a gender distinction in oblique third-person singular pronouns in Semnani (Pakhalina 1991, 191); many Tafreshi varieties, Abynei, the Abu Zeyd variety of Abādi, and Jowšaqāni (LeCoq 1989, 313, 318); Yazgulami (J. Payne 1989, 433); and Kurdish.

These observations are intended as examples of a general trend. Whether a description of an Iranian language is the rare modern full-length opus such as Selcan’s (1998) of Zaza or a brief sketch such as those cited above by LeCoq (1989) of lesser-known western Iranian languages, a greater degree of overspecification and structural elaboration is always immediately clear. Certainly the other languages parallel Persian in isolated cases; compare the absence of case marking in Shugni, or the fact that in most Kurdish dialects, suffixes for present and past stems are as isomorphic as in Persian (Joyce Blau 1989, 331). Kurdish even surpasses Persian in underspecification in that the plural verbal suffixes are invariant for person.

But in the general sense, while some varieties are closer to Persian overall in underspecification than others, none equal it. Even where Persian has developed overspecifications of its own, these are often paralleled in other Iranian languages, for example, the development of an indefinite marker (*cf.* 5.3). In a description of Persian grammar, a statement that the language has developed an elaborification distinguishing it from its sisters is virtually unknown, in contrast to the utter typicality of such statements in descriptions of its sisters.

To the best of my knowledge, a table in which the morphology in an Iranian language other than Persian was represented by an all-white column while Persian’s was flecked with black and gray would be impossible. Even if one fashioned an all-white column from one of the varieties most closely related to Persian and most threatened by it, thus paralleling it especially closely in reduction, Persian would only mirror it rather than surpass it in whichever features one chose. As Windfuhr (1987, 530) states it, in terms of morphology, Persian is “the most atypical Iranian language.”

## 5. Chronological Perspective

### 5.1. *Old Persian Becomes Abbreviated Persian*

Persian became the anomaly of Iranian between its Old and Middle stages. As noted in the introduction, Old Iranian is represented by Old Persian and Avestan, similar enough to be considered dialects of the same language. Late in Old Persian (after the reign of Xerxes), the case system begins breaking down (Schmitt [1989a, 60] terms these “hair-raising mistakes” [translation mine]). But the scenario at the Middle Iranian stage beginning in the 300s BC is strikingly different.

As of Middle Iranian, the Iranian languages divide clearly into western and eastern groups. The western group is a mere pair, Middle Persian and its close relative Parthian. In the former, case marking remains only in high-usage remnants; oblique marking used with family relatives and on an oblique first-person singular pronoun. Interrogative pronouns are only sparsely attested in the Old Persian corpus (Kent 1953, 69; Schmitt 1989a, 76), but Avestan shows that Old Iranian marked “who” and “what” richly for case; “who,” for example, for six cases (Bartholomae 1883, 104–5). By Middle Persian the “who” and “what” pronouns are no longer declined (Brunner 1977, 89–90). Grammatical gender has also disappeared in Middle Persian. The *-ān* plural marker is used only with obliques; dual marking is gone. Adjectival concord is used only in the plural in early Middle Persian documents, and is extinct later. The relative pronoun no longer declines, worn down to *ī* from a bisyllabic Old Persian precursor that inflected for three genders including a neuter form whose root was phonetically distinct (*taya* as opposed to masculine and feminine *haya* and *hayā*). The present and aorist verb forms have collapsed. Middle voice markers have disappeared. Imperative markers are now identical to present ones. (Data, unless otherwise indicated, is from Sundermann 1989a; and Khanlari 1979, 232).

Parthian is even somewhat more reduced structurally than Middle Persian. Hortative *hēb* is almost lost. Persian’s development of new markers such as the “accusative” *rā* (cf. 5.3) began in Middle Persian, but Parthian shows much less evidence of such emergences. Parthian sheds the *ī* relative marker by the 200s AD at the latest and allows the omission of relativizers *čē* and *kē* more than Middle Persian (Sundermann 1989b).

## 5.2. Eastern versus Western

Taken alone, this might appear to constitute “normal” language change, perhaps exemplifying the well-known tendency of Indo-European languages’ “drift” toward analyticity. But the eastern Middle Iranian languages make this harder to support, this especially given that documentation of more of these languages survives than of the Middle Western ones: most copiously, Sogdian, Saka (including Khotanese and the lesser-documented Tumshuq), Bactrian, and Khwarezmian. All of them display ample losses from Old Iranian, yet preserve a great deal more of its legacy than the western varieties, being “Germans” compared to Old Iranian’s Icelandic.

How “natural” was the leap from Old to Middle Persian, for example, when in the east, Sogdian light noun stems retained six cases and heavy ones two, with Old Iranian declensional distinctions contracted but still alive? All three genders survived (for both light and heavy stems), and the dual held on as a numerative form used with numbers, with separate allomorphs for masculine as opposed to feminine and neuter nouns. Adjectives remained fully inflected; relatives inflected to a degree. Middle voice marking survived partially. Demonstratives were well on their way to grammaticalizing as definite articles. Sogdian had the most overspecified verbal morphology in Middle Iranian, innovating new markings (i.e., irrealis) as old ones wore away (i.e., aorist) (Sims-Williams 1989b). Sogdian, in other words, was mutating normally, leaving no surprise that one of its descendants is today’s typically elaborated Yaghnobi.



Accent in Sogdian was not totally unrelated to survival of morphology. Early in its history a unique stress pattern placed the accent on final syllables in the absence of a long vowel or diphthong in the root. This ultimately created the distinction in noun stems between light and heavy, with the former preserving much more Old Iranian case marking than the latter (Sims-Williams 1989b, 182). Yet the fact remains that even the “heavy” stems retained gender distinction, and that inflectional elaboration remained so robust otherwise in the grammar.

Khotanese was similar, with six cases, simplified but robust declensional classes, three genders, retention of imperative, subjunctive, optative, and injunctive paradigms from Old Iranian, and innovations such as the redistribution of the active/middle voice marking distinction into one between transitive and intransitive verbs, and even new nominal declension classes (Emmerick 1989). Khotanese did exhibit signs of drift: dual marking was already marginal, case endings were less robust in later documents (ibid., 223), and phonetic erosion is obvious over time in the documentation: *hvatana-* “Khotanese” becomes *hvatāna-*, *hvaṇna-*, *hvana-*, and *hvamṇ-* (Emmerick 1998, 609). Yet even in its later attestations, idiosyncratic distinctions remained between the form of stem used with nominative singular versus other cases singular and plural, such as *hve*’ *hva’nd-* “man” (Emmerick 1989, 217). Western Middle Iranian had all but lost distinctions like this.

Bactrian is occasionally referred to as having “drifted” in a fashion similar to the western varieties by the first century (Sundermann 1989b, 111), but then this was still a language with case marking, regular plural marking, remnants of grammatical gender, a declineable relative marker (Sims-Williams 1989a), and other features unknown to Middle Persian and Parthian in the west. Khwarezmian was like its sisters, such as its having five cases, two genders, traces of dual marking, distinct pronoun forms for nominative, accusative, and cliticization, and a definite article (Humbach 1989).

In general, the eastern varieties were more conservative languages, with final syllables less fragile than in the west, and reflexes of the augmented imperfective retained (Old Persian aorist *adā*, imperfective *adadā* “created” [Kent 1953, 90]; Sogdian *ṣrāmāy*, *frāmāy* “to hear,” from an augment between preverb and root \**fra-a-māya-* [Sims-Williams 1989b, 187]).

### 5.3. *Middle Persian to Modern Persian*

The pathway from Middle Persian to Modern Persian between the 600s and 800s AD was a matter of a few more elisions, unremarkable in themselves and analyzable as natural symptoms of a grammar now founded upon an analytic homology, tending to shed features more typical of a synthetic one just as Modern English would have shed *whom* short of prescriptive influence.<sup>4</sup>

Middle Persian’s remnants of case and productive inflectional plural marking disappeared entirely. Ergativity in the past tense, already variable, disappeared (Sundermann 1989a, 153), last occurring in vestigial form in early Modern Persian largely restricted to the third person (Gernot Windfuhr 2004, personal communication).

Another likely symptom of the analytic tendency was the decreasing use of the causative inflection (*xor* “to eat,” *xorān* “to feed”) in favor of compounds consisting of a nominal and *kardan* “to do,” as in:

- (8) u-rā      bidār      kard.  
       he-ACC awoken did  
       ‘He woke him up.’ (Windfuhr 1987, 538)

This was a general symptom of a lexicon reduced to about 200 to 250 verb roots (Lazard 1989, 286) making heavy use of compounds combining nominals with a small subset of verbs including *kardan*, with *šodan* “to become” also especially central (*bidār šod* ‘he woke up’ [ibid., 538]). This compounding tendency is common in the Iranian languages, but no grammar I have consulted suggests that this trait has marginalized the causative inflection to the degree that it has in Persian (although Windfuhr 1987, 538, indicates that its use in Persian has increased recently due to increased linguistic consciousness among writers).

Similarly, Old Persian had a passive suffix: *bar-* “carry”; *abariya* “he is carried” (Kent 1953, 73). Middle Persian retained this as *-ih-*; Modern Persian shed it and conveys the passive periphrastically with the verb “become.” This contrasts with most other Iranian languages, which have either preserved the Old Persian marker (Tati *rije* “pours,” *rijie* “is poured” [Yar-Shater 1969, 246]) or innovated one (Kurdish *něj-ra-we* “buried” [Joyce Blau 1989, 332]).

Just as English developed *do*-support after the Viking invasions, Mandarin developed “transitive” marker *bǎ*, and the new Arabics conventionalized aspect prefixes, Modern Persian has developed local elaborifications. The overspecification of indefiniteness and definiteness in the noun phrase is a particularly salient example. The *i* clitic marks indefiniteness (*ketāb-i* “a, some book”; *ketāb-hā-i* “some books”); it began as Old Iranian’s *aiwa* “one,” and in Middle Persian was only used with singular nouns but in Modern Persian generalized to the plural (Windfuhr 1987, 541). The postposition *rā* functions not, as previously often analyzed by grammarians influenced by Latin and Greek, as an object marker but as a specificity marker, its function illuminated in contrasts such as:

- (9) (a) Xāne-i              āteš zadand.  
       house-INDEF they burn.PAST.3PL  
       ‘They burned a house.’  
       (b) Xāne-i-rā              āteš zadand.  
       house-INDEF-SPEC they burn.PAST.3PL  
       ‘They burned a certain house/they burned this house, you know.’  
       (Windfuhr 1987, 534)

This marker began in Old Persian as *rādi* “concerning” and was optionally marking indirect and direct objects in late Middle Persian; it conventionalized in its current function in early Modern Persian (Windfuhr 1987, 541).

There were also subtractive changes in early Modern Persian of the sort whose equivalents were occurring in its sisters as well, exemplifying the fact that ordinary language change entails both additions like *i* and *rā* as well as eliminations. In earliest Modern Persian, today’s imperfective *mi-* prefix, when used with verbs along with the *be-*verb in a perfect construction, connoted the continuous facet of that aspect, such that *mi-xarid-e ast* meant “he has been buying,” whereas in Modern Persian the prefix is restricted to the inferential past and the counterfactual. By the 1200s AD

the continuous meaning was restricted to the third person and is now extinct (Windfuhr 1979, 85). The prefix *bi-* in Old Persian was a perfective marker but is now restricted to the subjunctive and imperative (ibid., 95–96).

But remodelings of this kind are known to no grammar's mutation over time, an analogy being the eclipse in various regional Italian dialects of the marking of a subclass of unaccusatives with the use of a *be*-verb rather than a *have*-verb in perfect constructions, in favor of a new division of labor between *be* and *have* according to person, with *be* used in the first and second persons with all verbs, as in Eastern Abruzzese with transitive verb "to write": *sòšcrittu* "I have written," *sci šcrittu* "you have written," but *a šcrittu* "s/he has written" (Bentley and Eythórsson 2001, 64).

But the crucial disjunction had been between Old and Middle Persian, especially striking in that only three hundred years separate the last Old Persian documents and the first Middle Persian ones (Khanlari 1979, 231). Meanwhile, unattested western Middle Iranian varieties developed into those of today such as the saliently not "marvelously simple" Kurdish, Zaza, Gurani, and Tati, while varieties in the east, also mostly lost to history, developed into today's eastern varieties. All the modern varieties are well marked by the effects of retraction of the accent from final syllables—but none to remotely the degree in Persian.

## 6. What Happened to Persian?

### 6.1. *Temptations*

There are assorted potential explanations that can be dismissed. For example, despite the rich Arabic admixture in Persian's lexicon and the vast literature examining the cultural hybridization occasioned when Islam met Persian folkways, the arrival of Islam in Iran cannot have had any appreciable impact upon the structural contrast between Old and Modern Persian.

For one, the Arabs conquered the Sasanid Empire in 642 AD, much too late to effect a change clearly documented to have taken place nine hundred years before, by the fall of the Achaemenids to Alexander the Great. The wealth of Arabic lexicon, and even what can be analyzed as marginal Arabic-derived rules (e.g., Windfuhr 1979, 81–83), in Persian was the result of adstratal contact, but this took place without interfering with Persian grammar in any substantial way, just as in the impact of Norman French upon English. As always, in language contact, lexical mixture need not entail structural impact.

Even a case that Arabic contact was responsible for the minor reductions that Modern Persian displays compared to Middle Persian—the "finishing touches"—is extremely fragile. Documentation of the Arab presence in Iran for its first few centuries is very sparse, and even afterward, writers have all but nothing to say about ordinary people, among whom crucial language contact processes presumably would have been taking place (Cahen 1975, 305, 310). Nevertheless, it is recorded that 200,000 Arab soldiers and their dependents migrated to Iran in 672 and 673 AD—but to Khurasan (Zarrinkub 1975, 29), far northeast of the Persis region inhabited by Iran's rulers and from which most Persian writings emanated. Arabs preferred to migrate

to Khurasan and other regions such as Qum (far northward of Persis) and Sistan (far eastward), finding the climate more to their liking (Cahen 1975, 306–7). Arab migration to Persis itself appears to have been minor (Zarrinkub 1975, 27), and, overall, there was never a truly large-scale migration of Arabs into Iran.

Frye (1963, 113) surmises that mixture between Medes and Persians broke Old Persian down, but this again hastily equates contact with reduction. To attribute Persian's anomaly to population mixture alone would require that languages in other intensely hybridized areas like the Balkans, southern China, or sub-Saharan Africa (where Bantus met Khoi-San speakers) display similarly peculiar reduction compared to their relatives, whereas languages like Albanian, Hmong, and Xhosa obviously do not. Similarly, Sundermann (1989b, 110) supposes that Persian reached its state due to rampant mixture between speakers of closely related languages, but this leaves the question as to why even other Iranian languages did not suffer the same reduction: Pashto, for example, has developed amid similar mixture between sister languages.

## 6.2. *Persian under the Achaemenids*

We seek a historical factor both unique to Persian and predictable as a cause of structural reduction rather than just mixture. Here, it is crucial that 1) Old Persian was the language of a long-lived dynasty in Iran, and 2) Middle Persian appears on the record after the fall of this dynasty, the Achaemenids (550–330 BC). Our task, then, is to identify aspects of the use of Persian under the Achaemenids that would plausibly have streamlined its structure; namely, widespread second-language acquisition.

### 6.2.1. THE PLACE OF ARAMAIC

Meillet (1912) anticipated this analysis in supposing that Old Persian's decay was due to its use as a "Grossreichssprache" throughout the empire. Indeed, under the Achaemenids the Persian Empire extended through modern-day Turkey across to the western edge of the Black Sea in Greece, southward to modern-day Egypt, and eastward to modern-day India, as well as regions today contained in Afghanistan, Pakistan, Turkmenistan, Uzbekistan, and Tajikistan.

Yet, in point of fact, the Achaemenids felt almost counterintuitively little need to impose Old Persian upon their subjects. Despite observations such as Cook's (1985, 290) that the Achaemenids' rule "was one which induced racial and cultural fusion," all evidence confirms that the lingua franca under which this was accomplished was Aramaic. The standard practice was that documents and missives were dictated in a local language (including Old Persian), written by the scribe in Aramaic, and then read back to the recipient in the local language at the destination. In inscriptions and on seals, Old Persian is generally used alongside other languages. While, for example, Elamite (the language of the peoples who inhabited much of Iran before the Persians' arrival) was often used by itself, Old Persian almost never was (Cook 1983, 69; Root 1997, 232, 235). There is no evidence of Persian being spread across the empire as Latin was in the Roman Empire (Briant 1996, 525), such that Persian

today is not spoken across the former Persian Empire's expanse as Mandarin Chinese is spoken throughout China; Arabic throughout the former Islamic Empire; and daughters of Latin throughout Western Europe.

Szemerényi (1980) surmises that this use of Aramaic was the culprit in simplifying Old Persian. But legions of other languages in the empire submitting to Aramaic under the same conditions remained at a normal level of complexity (including Eastern Iranian languages). Official languages, especially those largely used in writing, do not regularly abbreviate the structure of languages lower in the diglossic context, as witnessed by, for example, most languages of Africa, spoken in countries where colonial European languages (or Swahili) occupy the sociological space similar to the one that Aramaic did under the Achaemenids. Yet there is nothing "streamlined" about Kikuyu or Bemba.

### 6.2.2. OLD PERSIAN IN IRAN

We are pointed, then, to the use of Old Persian in Iran itself, namely in the capitals (the Achaemenids were rather nomadic in this regard) where, presumably, language use would eventually determine what was felt to be its standard form and would be committed to writing.

Here, however, the historical sources are of only limited use. The majority of what is known about the Achaemenid reign was written by Greek and Roman scholars such as Herodotus, who not only wrote before the institution of academic standards of objectivity but also had little interest in the sociological and ethnological observations that fascinate scholars in our era. These writers were concerned largely with issues of warfare, court ritual, and dynastic succession, under a tradition that "frowned on detailed descriptions of far-away peoples and places" (Bivar 1983, 21). Beyond these accounts, scholars are left to rely largely on inscriptions—generally utilitarian and laconic—and the compact proclamations on coins and seals. There exists no body of literature written by Persians themselves to give us a more local and sociologically informed account of ordinary populations in Iran of the period.

Among the fragments of useful information on immigration in Persis that have survived, however, are descriptions of robust interethnic mixture. King Darius I himself described in an inscription a multiethnic crew—including Babylonians, Ionians, Lydians, Egyptians, and Medes—constructing and decorating buildings in principal Achaemenid capital Susa, with Herodotus noting that rebellious subjects tended to be brought to Susa for such activities (Frye 1963, 96). Eratosthenes summed up Persis as "Iranians living among barbarians" (Wiesehöfer 1996, 78). The precious Fortification and Treasury tablets, discovered in the 1930s, depict urban Persis at the turn of the 500s BC to the 400s BC. They describe workers as mostly foreign, some enslaved from conquests. Well over five hundred Egyptians at a time worked as stonemasons, while there were also Indians, Eastern Iranians, Babylonians, Cappadocians, Carians, Greeks, and perhaps Lycians and Arabs, as well as many hundreds of "Ishkuduru" perhaps from Thrace. Some of these workers were living in Persis only on temporary assignment, but then parties of workers arriving were typically two and

a half times the size of those leaving, and seem to have often included whole families (Cook 1983, 87).

A spot depiction from Babylonia under the Achaemenids is also pertinent in demonstrating the ethnic mixture typical under the empire: Egyptians, Jews, Moabites, Afghans, Indians, West Semites, Elamites, Lydians, Greeks, Phrygians, Carians, and Arabs lived together, and it was described that “a contract of marriage between a Persian and an Egyptian woman is written in cuneiform script. Arameans, Babylonians, Egyptians are witnesses” (Cook 1985, 290 ff.).

This latter description is useful only in depicting interethnic mixture rather than the use of Old Persian, because presumably the language of daily communication in Babylonia under the Persian Empire was not Old Persian. Sources leave a question as to just which language this was, and, in fact, there is no concrete indication even of what language the workers in Persis were using. However, it would seem likely that in Persis itself, this would have been indeed Old Persian. The formal use of Aramaic in writing, chosen because of its familiarity across the vastness of the empire, need not signal that the Persians went so far as to withhold their language from newcomers residing in their ethnic heartland. Such would have been distinctly odd given that the Persians of the period, typically of rulers of empires, were known for their ethnic pride (e.g., Briant 1996, 194).

In general, the evidence left behind by written documentation can be starkly misrepresentative of daily oral linguistic reality. If writing and scholarship were as restricted in domain today worldwide as they were twenty-five hundred years ago, then a hypothetical scenario twenty-five hundred years from now, under which most of the world’s languages had been replaced by English, is illustrative. Analysts might well miss that Cantonese rather than Mandarin was once spoken in Guangzhou province, or that new Arabics rather than Modern Standard Arabic were spoken in the Arab world.

A useful illustration is that Fanakalo, a pidginized Zulu, was the *lingua franca* among blacks as well as between blacks and whites in the mines of Windhoek, Namibia, in the first half of the twentieth century. Yet in thirty years of archival records from 1920 to 1950, there is not a single mention of Fanakalo: there was no reason for anyone besides a scholar to acknowledge it in writing (Ferraz 1984, 110).

Scholars, such as they were, in the Persian Empire left us no accounts of what they experienced. From an empire where writing was typically confined to stone monuments, official missives, and inscriptions on coins, and was addressed exclusively to a cross-continental audience to whom Old Persian was largely unknown, we would be surprised to encounter any evidence at all of everyday linguistic reality in Persis itself.

### 6.2.3. ELAMITES

An additional factor in the reduction of Old Persian may also have been acquisition by the peoples inhabiting most of the Iranian plateau before the arrival of Iranians, the Elamites. Elamite (of unclear relationship to other languages but likely related to Dravidian) was an official language under the Achaemenids, regularly used in

written communications. Elamite writings often have many loanwords and syntactic calques from Old Persian (Frye 1963, 70; De Miroschedji 1985, 302), suggesting contact between the languages which could well have operated in the opposite direction as well.

Indo-European speakers began making brief appearances in Iran as early as the second millennium BC, and the Persians and Medes are first reported in the 800s BC. This alone suggests the gradual kind of population merging which generally occurs without significant effect upon the transmission of grammatical complexity. However, under the Achaemenids in the 500s BC there was a massive influx of Iranians, and this was just as Elamites were changing from a nomadic to a sedentarized lifestyle (*ibid.*, 294–95). De Miroschedji (*ibid.*, 304) characterizes Persian culture as originating in an absorption of a large population of Elamites into an Iranian population.

### 6.3. *Parthian*

A remaining issue is that Middle Persian's close sister Western Iranian language Parthian was as equally reduced from Old Iranian. Here, an explanation appealing to non-native acquisition under empire points to the fact that Parthian was the language of rulers of Iran under the Parthian Empire of 247 BC to 224 AD.

One source of Parthian's reduction may have been that the Parthian Empire began with the overthrow of the Seleucids of the post-Achaemenid Greek interregnum by the Parni tribe, coming from the region between the Caspian and Aral seas. Taking over the Parthian province, the Parni adopted the name of the indigenous Iranian group who were the source of the province's name. The Parni, who founded the Arsacid dynasty, would have spoken a now lost Middle Iranian language usually thought to have been related to Scythian (described by Justin as akin to this and Median). However, the language they used as rulers was Parthian, closely related to Middle Persian. The conquerors not only adopted the name of the Parthians but also their language.

We could reconstruct, then, that one factor in Parthian's reduction was that the rulers acquired a language non-natively and then, as the dominant class, were in a position to commit the language to writing. Because the language had not been written before, prescriptive norms would have exerted less influence upon its scribes than normal, the result being a reduced version of a nonnatively acquired language, enshrined in official documentation for the ages. Perhaps this would explain Sundermann's (1989b, 110) otherwise rather mysterious observation that Parthian was less tightly conventionalized than Middle Persian in terms of word order and obligatoriness of the expression of various markers. Might the records of Parthian represent, to an extent, an incompletely mastered language?<sup>5</sup>

We must also look, however, to the possibility of widespread second-language acquisition of Parthian beyond the Parni. Sadly, our data on populations under the Parthians are as scanty as those under the Achaemenids, the historical account largely written by Greeks and Romans at war with the kingdom and reporting in distant scorn, little interested in the peoples inhabiting the region (Bivar 1983, 21). However, lists of names suggest that Parthian communities were "extremely mixed" racially (Colledge 1967, 86–88): the names of the rulers of the Characene kingdom (atop



the Persian Gulf) are variously Greek or of various eastern origins; Dura Europos (in the Fertile Crescent) reported names Greek, Semitic, and Iranian. Aramaic speakers were a dominant presence in western Parthia. Babylonians (speaking Akkadian) also inhabited the empire in large numbers. Greeks, remaining after Alexander the Great's victories of yore, were especially dominant in many towns until the first century AD.

While we are dealing here in speculations based on fragmentary evidence, the difference in degree of reduction between Middle Persian and Parthian in the west, and the Middle Iranian varieties in the east is so striking that this account would seem worth consideration.

#### 6.4. *The Eastern States*

A remaining issue is that many of the eastern Middle Iranian languages were the lingua francas of kingdoms in the areas that are now parts of Uzbekistan, Tajikistan, and Afghanistan (Khotanese was the variety of Saka spoken in one city, Khotan). These kingdoms were variously subdued by other powers such as the Achaemenids, the Greeks, the Sasanids, Arabs, and peoples of China. A thorough address of the Persian question requires considering whether there were large-scale incursions of immigrants into Sogdiana, Bactria, or Khwarezmia. If there were, then the elaborated character of Sogdian, Bactrian, and Khwarezmian contradicts our hypothesis.

However, data on life in these places at the relevant time is too sparse to be of any real use. Once again, most of what is available until the 600s AD is passing impressions by foreigners, in this case not only Greeks and Romans but also Chinese to the east of them. But here the data are even scantier than for Persis or Parthia; even basic facts such as whether Bactria ruled Sogdiana for a period are unclear in the record. Scholars have been left to piece together a dim, fragmentary picture of the antiquity of the "Transoxiana" region by combining brief observations by travelers with ingenious but broad extrapolations from coin inscriptions and remnants of irrigation technology.

But what we do know suggests that at the point when we encounter the languages in documents—the early 300s AD for Sogdian, the 100s AD for Khwarezmian, and around the same time for Bactrian—the kingdoms were ruled by nomadic peoples, these only settling in large numbers starting in the 200s AD. The Greeks, for example, appear never to have controlled the areas in anything approaching their entirety. Historiographical and archaeological data suggest that the kingdoms were loose confederations, with small cities few in number (Zeimal 1983). These were less empires than decentralized regimes of only moderately advanced development, at most times under the yoke of larger powers. This would presumably have made unlikely the large-scale importation of foreigners and especially heavy migration to the kingdoms' cities by rural people.

Given that they were used as vehicular languages, it is almost certain that there were thriving nonnative varieties of Sogdian, Bactrian, and Khwarezmian. However, the elaborated character of their documentation suggests that approximate acquisition was never widespread enough to affect the standard varieties. This is especially so since their elaboration persists over time. An analogous case would be Fula, spo-



ken in a wide range of West African countries mostly as a second-language variety. Second-language speakers use a significantly reduced rendition of the language, especially predictable given its vast array of noun classes rendered with fiercely complex consonant mutations. But this rendition only happens to have been described, briefly, by LaCroix (1967). First-language speakers of Fula retain the full variety that one learns of in the language's grammars.

## 7. Genetic Evidence

Certainly, the historical evidence presented above qualifies only as glimmers of a complete account. These snapshots may well depict a general reality of the period, like fragments of a silent film mostly decomposed. Yet just as we can only be but so satisfied by hypothetical reconstructions of the entire film from which such fragments survive, Darius's boastings and passing comments by foreign war correspondents can comprise only the beginnings of a solid historical explanation of Persian's reduction. As to Parthian, even the precise nature of the Parni takeover and its participants is vastly unclear in the fragmentary and allusive sources, long subject to competing interpretations. We miss population figures, substantial material on the languages at graduated intervals of time, more precise identification of the population's origins, observations as to what languages they spoke natively, and what languages they used with one another, and so on.

All indications are, however, that these facts are lost to history. They occurred in a world in which writing was a marginal and official activity largely confined to the monumental, liturgical, and administrative, in which no one was interested in committing to writing nugatory observations of humble commoners, and certainly not particulars of the nature of their rendition of Old Persian, a language most of the observers did not even speak. Archaeological ruins are mute as to what languages their inhabitants spoke. The human fossil gives no more attestation of language than a stone.

Or does it?

Recent developments in genetic research trace early human migrations with a precision almost inconceivable until recently, referring to background mutations on the Y chromosome as they vary among population groups. Over time, small, harmless mutations accumulate on human chromosomes within populations and survive across millennia. Tabulated through sampling of peoples throughout the world, these mutations serve as markers that allow geneticists to trace the movements of early humans across the planet. Often, these mutations correlate with the distribution of linguistic families, this reinforcing the increasingly detailed reconstructions of population movements that genetic research allows (cf. Cavalli-Sforza and Cavalli-Sforza 1995).

The Y chromosome marker termed M17 appears to correlate with the spread of Indo-European into Asia, dating to ten to fifteen thousand years ago and with the most diversity among variants in southern Russia and the Ukraine, corresponding with the hypothesis that Indo-European originated among the Kurgan peoples in that region. M17 is rare among Middle Easterners, but in India is robust among Indo-

Aryan speakers—but much less so among Dravidian speakers. This correlates with archaeological data in indicating that Indo-European was brought from today's Ukraine to India, bypassing the Semitophone Fertile Crescent to the southwest.

Of course, that pathway to India—or the even one posited by Renfrew (1987) beginning from present-day Turkey south of the Black Sea—would have been intermediated by a movement through Iran. But there is an intriguingly specific kink in the data: M17 is intriguingly low in none other than western Iran, while in eastern Iran it is virtually as robust as among Indo-Aryan speakers (Wells et al. 2001, 5; Wells 2002, 167). For Wells and his colleagues, this qualifies as a quizzical but minor curiosity, for which they can offer only random speculations. They offer three.

First, Wells surmises that when Iranian speakers migrated to the western region, there was already so great a population density that Iranian speakers imposed their language via cultural dominance despite being a numerical minority. But this ultimately is but a guess, designed as a potential explanation for a peculiar incompatibility between genes and language distribution. Geneticists such as Cavalli-Sforza have plausibly applied this “elite dominance” explanation to the movement of Indo-Europeans westward into Europe, supported by ample archaeological and linguistic evidence that other peoples had long before occupied Europe (the linguistic isolate Basque in France and Spain being a prime demonstration in the linguistic vein). But appealing to this model to explain a genetic discrepancy between western and eastern Iran requires evidence that population density was greater in the west before the Iranians came.

This evidence, however, does not exist. Because the peoples in the west were observed and written about by reporters from the literate West, there is a great deal more data on the Elamites, Kassites, and other autochthonous groups of the area than of peoples elsewhere in what is now Iran. There is, however, archaeological evidence of pre-Iranian inhabitants in eastern Iran (Diakonoff 1985, 54; Daniel 2001, 24). Given that today's political boundaries did not exist when the Iranians arrived (generally assumed as between 3000 and 1000 BC), it is also relevant that there were peoples established in areas adjacent to modern-day eastern Iran. This includes Uzbekistan, Turkmenistan (this last the location of the Jeitun and Namazga-Tepe cultures [Diakonoff 1985, 54; Sinor 1998, 715K]), and western Afghanistan (Srivastava 1996, 2–4). Thus there were human inhabitants both west and east, and geneticists are faced, *tout court*, with a present-day anomaly for which the historical data offer no correlation.

Second, Wells surmises that “an Indo-Iranian language may have been the *lingua franca* of the steppe nomad and the surrounding populations,” and “over time, this language could have become the predominant language in Persia” (Wells et al. 2001, 5). If I read him correctly here, this would mean that an Indo-European language was already spoken in western Iran before the descendants of the Kurgans arrived. But this would seem to put into question the centrality of M17 to the dissemination of Indo-European, otherwise so neatly apparent in terms of genetics, language, and archaeology.

Third, Wells guesses that the Dasht-e Kavīr and Dasht-e Lūt deserts of central Iran would have reduced the number of Indo-European-speaking migrants who moved to western Iran. But this implies that humans necessarily would have moved

first to eastern Iran and then would be confronted with the desert barrier to westward movement. But whether western Iranians had arrived from the east is a matter of considerable controversy among specialists in Iranian archaeology, partly because of a simple question as to what would have blocked early Indo-European speakers migrating from the north from spreading down to the west of the deserts as well as to the east of them. No one viewing a map of the region and unaware of Wells's hypothesis would independently suppose that migration to western Iran would have been blocked, or would have been even less likely than migration to eastern Iran.

Here, it is crucial that so many of the peoples documentation designates as migrants to Achaemenid Persia were from, specifically, the Middle East where M17 is so weakly attested: Babylonians, Egyptians, and various Semitic peoples. For the geneticists, the sparseness of M17 in western Iran is a mere statistical wrinkle. But in terms of the language contact scenario, it may signal heavy admixture from populations within which M17 was weakly represented. That is, the otherwise baffling sparseness of M17 in precisely the portion of Iran within which we are faced with a peculiarly high degree of grammatical streamlining can be taken as genetic evidence of heavy population movement into that region, only glancingly documented by contemporary observers.

This would mean that Indo-European speakers indeed migrated into Iran west of the deserts as well as east, but once settled in that westward region, later under the Achaemenids saw a flood of migration by peoples from the Middle East. These people would have been much less likely to have the M17 gene, and left that legacy to peoples of western Iran today. But for our purposes, it is more important that they would have spoken languages distinct from Old Persian, and would have acquired and passed it on in an abbreviated rendition. Here, it may be germane that skeletal remains from the former Parthian Empire have been reported to contain a mixture of specimens of Iranian, Greek, and eastern stock (Debevoise 1938, 1).

## 8. Conclusion

The keystone observation in the case presented in this chapter is that Persian is uniquely abbreviated morphologically amid a vast array of sister languages, to a degree that strongly suggests that something irregular affected its development.

Evaluation of the argument cannot be predicated upon a possible sense that to be valid, the case would refer to documented incremental changes in Old Persian over time correlated closely with chronological demographic data from ancient Persis. This is not because data of this kind are not awaiting discovery, but rather, do not exist. The period during which Persian acquired its peculiar character was in distant antiquity, centuries before the birth of Jesus, when for Persians the sheer act of writing was largely ceremonial and administrative, conducted almost exclusively in foreign languages comprehensible to subjects who mostly did not speak Persian. Inevitably, sociolinguistic or demographic data of any substance from these ancient realms, in which language was largely oral and our modern conception of social science was alien, are forever lost to us.

However, this acknowledged, the anomalous nature of Modern Persian remains. And rather than being a tongue with a history unextraordinary, Modern Persian is the descendant of the native language of rulers of the vastest empire ever ruled by speakers of an Iranian variety, this empire documented as having been a vibrantly multiethnic one. How likely is it that there is no connection between this and the language's qualitative contrast with its relatives?

Certainly, one might choose that analysis—that is, assume that Old Persian's baroque elaborification melted away to such a striking extent in just this one of its descendants solely by chance. Such an approach is hardly unscientific. Many prefer to view the emergence of life itself as a fortuitous result of chemical conditions on our particular planet, with this mere roll of the dice being responsible for us humans' having emerged with the ability to ponder the issue alone in the universe.

But I suggest that Persian's erstwhile status as a language of empire renders a contingent analysis of this kind unnecessary, and even uneconomical. While all grammars abbreviate their ancestors' machinery over time, there exists no Slavic, Uralic, or Algonquian analogue to Persian in comparison to its sisters. No Slavic language excuses the learner from mastering exquisitely elaborate morphological paradigms; no Uralic language has mysteriously shed the family's famous spatial-marking suffixes; there is no bracingly analytic Algonquian language that does without obviative marking.

Persian, then, is as peculiar as a Chinese language without tones. I argue that this nothing less than requires an interruption in its transmission at some point in its history, and that the gulf between Old and Middle Persian drives us to reconstruct this as having occurred under the Achaemenids. With it agreed at the very least that nonnative acquisition abbreviates natural language grammars—most unequivocally demonstrated by the clinal extreme of pidginization—the fact that Persis under the Achaemenids' rule is documented to have been a distinctly heterogenous society must be received as crucial and decisive information.

I further submit that the arguments in the previous chapters indicate that a stark contrast in elaborification between a language and its sisters is not just sometimes but always due to extensive nonnative acquisition. The peculiarity of Persian, then, can be seen as positive evidence of heavy immigration into Achaemenid Persia, a perspective logically justified by the fact that when many other languages contrast similarly with their sisters, sociohistorical explanations are more conclusively correlative with the anomaly, as in the Chinese, Arabic, and Austronesian cases.

Thus, to be sure, the concrete evidence of extensive nonnative acquisition of Old Persian in Persis is suggestive at best. But in evaluating this chapter's argument, we must return always to the simple fact of Persian's anomalous character. Not a single other Iranian language, be it an obscure tongue spoken by a few hundred or a widespread lingua franca with a written literature, parallels Persian in its streamlined nature. If Persian acquired its modern character via mere "drift" driven by the withdrawal of stress from its final syllables, then why not more Iranian languages in which this has happened *to the extent that it has in Persian*? Why not at least two or three? Why not just one?

In the absence of even one Iranian language akin to Persian in this regard, it would seem unavoidable to suppose that something happened to Persian that did

not happen to its sisters. The linguistic, historical, and perhaps even genetic evidence combine to indicate that what happened to Persian was the moderate abbreviation familiar to most of us from our own and other students' diligent but imperfect renditions of foreign languages practiced in the classroom or abroad. Modern Persian represents what happens to an early Indo-European language when acquired by learners past the period when they are capable, or even desirous, of acquiring new languages perfectly.

# Colloquial Arabic

## *A Mysterious “Drift”*

### 1. Introduction

Arabic began as a tongue spoken mostly by bands of nomads in Arabia, little known beyond its confines. But when Mohammed brought speakers of this language together under a new, aggressive Islamic Empire in 630 AD, Arabic met the world. The earliest records of Classical Arabic, the nomads’ poetry and the language of the Qur’ān, reveal a language with a rich nonconcatenative morphology, affixal marking of case, definiteness, two moods, tense-sensitive negation strategies, and quirky lapses of iconicity in gender marking such as assigning feminine singular concord to inanimate plurals and marking masculine nouns with feminine numeral forms and feminine nouns with masculine ones.

The Islamic Empire conquered peoples as far north as Syria and Iraq, as far east as the Indus River, as far west as Morocco and Spain, and southward to Chad and Sudan. This spread was quite rapid, and in each new location, Arabs encountered speakers of other languages, many of whom had urgent reason to acquire the language quickly, especially since it was the vehicle of a successfully imposed religion.

Based on the previous chapters, we would predict that the result here would be that a millennium and a half later, the Arabic spoken natively in these locations would be markedly streamlined in comparison to the language the nomads of the seventh century had spoken. Hardly would we expect pidgins or creoles, but readily we would expect “Mandarin” versions of the language of the Qur’ān. As the reader has good reason to expect, this chapter will argue that this is exactly what happened to Arabic across the Islamic Empire.

### 2. The New Arabics: Business as Usual?

The qualitative difference between Classical Arabic and the modern Arabic “dialects” is readily apparent, but debates and scholarship on Arabic diachrony and dialectology

tend to be predicated on a tacit assumption that the difference is an unremarkable manifestation of how grammars normally change over time.

This conclusion is understandable. Moroccan Arabic, with its weakenings and elisions of short vowels, elimination of Classical Arabic's case markers, marginalization of several of its paradigms of nonconcatenative verbal morphology, tendency to replace or reinforce these with affixes and particles, and replacement of Classical Arabic's VSO tendency with an SVO one, easily recalls the difference between French and Latin. Modern linguistics was created by Indo-European-speaking Westerners, most of whom were native speakers of Romance and Germanic languages. Because the Italic and Germanic subfamilies and the broad outlines of their historical development are so familiar, naturally the development from Latin to French or Old English to English will be processed as ordinary demonstrations of how grammars change. This is especially so since these scenarios are reconstructable from the concrete medium of writing, rather than depending largely on hypothetical reconstruction as does historical work on most of the world's six thousand languages.

In this light, we would expect that to many analysts, the differences between regional Arabics and Classical Arabic will appear to be business as usual. Claims that there was any appreciable interruption in transmission will appear unnecessary, insufficiently attendant to how "ordinary" sound erosion, morphological simplification, and reanalysis are. Thus the debate that Ferguson (1959) sparked with his argument that all of the new Arabics stem from a single koine was one about shared idiosyncrasies rather than reduction. Responses such as those from Blanc (1970), Kaye (1976), and Joshua Blau (1977) focused on whether the features Ferguson referred to suggested that the new Arabics all stemmed from a new rendition of Classical Arabic rather than Classical Arabic itself. However, the fact that over half of Ferguson's features were simplifications (affixal losses, phonetic erosions, paradigmatic simplifications) was treated as incidental, these developments seen as bread-and-butter commonplaces of how languages "drift" over time (e.g., Cohen 1970).

Many analysts will smoke out evidence of the changes as far back as Classical Arabic itself to argue against the changes having been sparked by later circumstances. Hence various scholars' proposals that Arabic had already lost its case markers before the Islamic period (Corriente 1971; Diem 1973; Zwettler 1978). In work on the historical development of the new Arabics, possible sources of transfer will be attended to more than structural reduction. Consequently, one of the main things that strikes Diem (1978) about the new Arabics is how little they have transferred from other languages rather than their reductive character overall, proposing that transfer was stanchied because Islamic cultural identity and the influence of the Qur'an led native speakers to resist foreign incursions.

The main analysis focusing on the reductive aspect of the new Arabics is Versteegh's (1984) monograph, noting that the development of the new Arabics parallels pidginization and creolization. However, Versteegh's conception would seem to have had little impact on subsequent inquiry into the subject (with Kusters 2003 a significant exception, to be referred to extensively later in this chapter). Moreover, Versteegh's book also addresses the new Arabics selectively, illustrating given points with passing reference to a single variety or a few of them. But while the

Arabist's background familiarity with all of the dialects allows them to assess the validity of Versteegh's arguments, the ideal presentation would address all of the dialects systematically, for the purposes of the linguist less familiar with Arabic and Semitic.

This chapter will complement Versteegh's presentation for the purposes of the overriding thesis of this monograph. I will second Versteegh and Kusters in arguing that the new Arabics represent conventionalized non-native renditions of Classical Arabic, rather than uninterrupted continuations of a parent language.

### 3. Arabic's Close Shave

#### 3.1. *Why So Much Loss?*

The difference between any modern Arabic variety and Classical Arabic is overwhelmingly one of subtraction rather than addition or elaboration. This contrasts with the difference between, for example, Slavic languages with Proto-Indo-European. The Slavic group has certainly collapsed some vocalic distinctions, shed some cases and lost some tense distinctions. These are typical processes of language change. But as often as not, and just as typically of language change, they have developed many new segments and phonemic contrasts of their own, and the aspectual systems they have developed amply "compensate" for the loss of tense distinctions. Overall, languages like Russian, Polish, and Serbo-Croatian do not give the impression of being abbreviations of Proto-Indo-European to any significant degree. Rather, as one grapples with the maddening complexity of Polish declensional morphology or the behavior of clitics in Serbo-Croatian, one is faced with modern renditions of Proto-Indo-European qualitatively indistinguishable from it in overall complexity.

This is typical of language change. The language that displays no losses, mergers, and collapses over time is unknown. But these processes are accompanied by innovations marking new distinctions, such as via "exaptation" of old material as described by Lass (1990), or the splitting of some cases into ones marking more specific concepts in the evolution of Samoyedic as described by Marcantonio (2002, 209).

In contrast, descriptions of modern Arabics are uniform in their general theme of decadence. Grammars of these languages are replete with observations that a given feature in Classical Arabic is eliminated or vastly simplified. Meanwhile, elaborification is only rarely encountered, and never of any extreme extent, such as the development of new allomorphic inflectional paradigms. An explicitly comparative description such as Altoma's (1969) of Iraqi Arabic compared to Classical Arabic unavoidably carries an impression of documenting a kind of "Junior Arabic," in comparison to Cantonese or Min's appearing to be "hyper-Chinese" compared to Old Chinese.

#### 3.2. *Clinal Rather Than Polar*

This is hardly to imply that the new Arabics have reduced Classical Arabic to the point of pidginization or creolization. For example, all of these dialects retain the notoriously arbitrary Arabic broken plurals for most nouns, formed by vocalic



alternations around the consonantal root, occurring in several patterns the appropriate one of which must be learned by rote for each noun; for example, *bayt* “house,” *buyūt* “houses,” *kitāb* “book,” *kutub* “books” in Gulf Arabic as in Classical Arabic.

Moreover, the regional Arabics have by no means shed all but a sliver of their original verb morphology as English did. All of them retain the elaborate system of verb measures, in which a triconsonantal root is inflected in the perfect and imperfect, and then subjected to an array of derivational alterations to encode valence distinctions, with this system then further complexified by exceptional processes that occur according to whether the second and third consonants of the root are identical, or whether the “weak” consonants [w], [j] or [ʔ] are first, second, or third root consonants. Thus in Classical Arabic’s Measure One, the perfect is *kataba* “he wrote,” while the imperfect is *yaktubu* “he writes.” In the perfect, Measure Three marks the reciprocal by lengthening the first vowel, and thus *kātaba* “he corresponded with.” Measure Eight has devolved beyond predictability but usually renders an intransitive meaning, such that *kataba* becomes via a prefix and infix *ʔiktataba* “he was registered,” while Measure Ten tends to connote request, and thus via prefixation and elimination of the first vowel, *kataba* becomes *ʔistaktaba* “he asked someone to write.” Although simplified to varying extents, this basic structure is vibrantly alive in the new Arabics, complete with departures from the basic patterns ranging from systematic exceptions to outright irregularities.

Thus our interest here is not in polar extremes, but in degree. The thoroughly reconstituted Arabic exists as Nubi Creole Arabic in Uganda and Kenya, where all of the above distinctions and more are lost. The regional Arabics represent a much more intermediate degree of reduction of Classical Arabic structure.

### 3.3. *The Data*

Nevertheless, that degree is more robust and, importantly, peculiar than many analysts’ work would imply. The figure below demonstrates the extent in question.<sup>1</sup>

Particularly important is the marked contrast between the Bedouin Arabic varieties, here represented by Najdi Arabic, and the others. The conservativity of the *nomadic* as opposed to *sedentary* Arabics is well known both to Arabist scholars and as a folk assumption among ordinary Arabs, who have been given since the birth of the Islamic Empire to designating Bedouins as retaining the language of the Qur’ān. The latter perception is highly exaggerated, based on Bedouin varieties’ conserving certain features such as an abbreviated rendition of Classical Arabic’s case and definite markers. However, the Bedouin varieties indeed retain much more of Classical Arabic’s complexity than the others, and this is crucial in that Bedouin communities have been much more sociologically isolated than the sedentary ones, which presumably has occasioned less non-native acquisition, and consequently less structural reduction.

It must be acknowledged that there is considerable variation within each variety, according to the urban/rural dimension, sometimes the Muslim/Christian/Jewish one, as well as sheer geographic variation across vast areas. Data from each in the table rep-

resent the varieties chronicled by the author of the source. Nevertheless, there is, to my knowledge, no dialect of new Arabic without the massive losses shown in the figure (nor has any writer I have consulted mentioned one); nor does there seem to exist a sedentary variety that parallels the Bedouin ones in moderate degree of loss.

I have selected the losses in the figure on the basis of those commonly noted in the Arabist dialectologist literature as significant in distinguishing Classical Arabic from all or most of the new ones, especially the features treated in Ferguson (1959), Cohen (1970), Fischer and Jastrow (1980), and Versteegh (1984). From the sum of those features, I have not included in the figure reductions of a type easily predictable as occurring even in several descendants of a language due to universal tendencies. This includes virtually inevitable developments such as that of denominal adjectival suffix *-iyy* to *-ii*, or the collapse of certain exceptional verb classes such as verbs with [w] versus [y] as the third consonant of their roots in Measure Three, an elimination of a fragile and specific distinction also dissolved elsewhere in Semitic such as in Hebrew and Syriac (Eastern Aramaic) (Kaye and Rosenhouse 1997, 297).

Since our interest is in losses, I have also not included features the above authors list that are solely differentiations, rather than losses, from Classical Arabic, such as Ferguson's observation that the new Arabics all have *šāf* rather than *raʿā* for "to see," or Cohen's that in one lesser-used broken plural pattern based on not three but four consonants (quadriliteral), new Arabics eliminate a first vowel [a] that was variable in Classical Arabic and shorten the third vowel (to the Arabist, the *ʃ(a)ʕaaliil* pattern becomes *ʃʕaalil*).<sup>2</sup>

Most striking is how many fewer black squares there are for Najdi. The other new Arabics differ in the extent of their losses: Nigerian is an extreme while Gulf has somewhat less loss, likely because Arabia is the Bedouin heartland. But the overall contrast is striking. As usual in this monograph, my research has not revealed to me an array of additional features for which a chart could be composed in which the contrast between Bedouin varieties and the others would be reversed. Moreover, as will be clear from remarks later in the chapter, if the number of features in the chart were expanded, the uniformity of the degree of loss among the new Arabics would lessen somewhat, but remain extreme to a degree demanding explanation.

### 3.4. Discussion of Features

*Interdentals.* Given the cross-linguistic markedness of interdental consonants, and the tendency for second-language acquirers unfamiliar with them to replace them with stops or other fricatives, the evanescence of Classical Arabic's *θ* and *ð* in so many new Arabics can be seen as evidence of an interruption in transmission. (Also, varieties with *ð* do not retain an original segment, but have developed this secondarily as a simplification of Classical Arabic's especially marked emphatic segment *dād*, likely pronounced as a lateralized dental [ɗl] [e.g., Kaye and Rosenhouse 1997, 275].) Predictably, influence from the standard can link interdental usage to register: in Gulf Arabic, educated speech tends to include interdentals while less educated speakers often replace them with [t] and [d], respectively (Holes 1990, 262–63). Meanwhile, in Syrian, interdentals are mainly used in some rural dialects (Cowell 1964, 3).

[illegible]

tense-blind negation										
no feminine comparative										
analytic comparative <sup>3</sup>										
undeclared rel. marker										
bimorphemic q-words										
loss of internal passive										
loss of subjunc. and SVO										
jussive										
verb patterns retained of 10	10	9	9	9	8	7	6	9	8	8

FIGURE 7.1. Losses from Classical Arabic in the New Arabics

*Weak glottal stop.* The voiceless laryngeal glottal stop was a central phoneme in Classical Arabic, distinguishing minimal lexical pairs and serving robustly as a trilateral root consonant. Even in Classical Arabic the typical fragility of the consonant was occasionally evident: “he will ask” could be either *yasʔalu* or *yasalu* (Kaye and Rosenhouse 1997, 277). But in regional Arabics, its phonemic status tends to be weak to nonexistent, often replaced in roots by glides or compensatory lengthening, or simply eliminated (Classical *badaʔa* “he began,” Egyptian *bada*), and restricted largely to explicit pronunciation or “higher” registers influenced by the standard. Egyptian is typical in tending to retain it in only high-usage forms such as the definite article *ʔil*, pronouns such as *ʔana* “I,” interrogatives such as *ʔēh* “what,” and functionally central morphological manifestations such as the first-person singular imperfect (*ʔaktib* “I write”) (T. F. Mitchell 1956, 115).

*Short vowel reduction.* Sedentary Arabics tend to eliminate short vowels in unstressed open syllables, such that Egyptian *kātib* “writer” becomes *katba* when rendered feminine. In most varieties the process is variable yet rampant, the underlying short vowels rendered mostly in the most explicit or prescriptive speech. This entails a contraction of the robustness of the distinction between short and long vowels in Classical Arabic in favor of consonant clusters (these in themselves not designable as “complex”).

In Bedouin varieties the phenomenon is less advanced, constrained by particular phonological environments rather than applying across the board (e.g., Ingham 1994, 16–19).

*/i/-/u/ collapse.* Of the three short vowels in Classical Arabic, /a/ remains relatively stable in the new Arabics while /i/ and /u/ are buffeted toward the central and midregions of articulation according to environment. The latter two are often tinted toward each other to the point of virtual identity as schwa on the surface.

*Case markers.* Classical Arabic marked case and definiteness on nouns with suffixes, as with *bayt* “house”:

	nominative	accusative/ oblique	genitive
definite	<i>bayt-u</i>	<i>bayt-a</i>	<i>bayt-i</i>
indefinite	<i>bayt-un</i>	<i>bayt-an</i>	<i>bayt-in</i>

These distinctions have disappeared in all of the sedentary Arabics, fossilized in set expressions and adverbial forms such as *dāyman* “always.” Genitive marking is retained via the analytic possessive and oblique marking via prepositions, but accusative is left to context.

Only Bedouin dialects retain the case marking in anything approaching productivity, albeit in collapsed, variable *-Vn* form insensitive to case, and mainly with nominals modified by adjectives (cf. Kaye and Rosenhouse 1997, 299).

*Feminine singular with inanimate plurals.* A notorious quirk of Classical Arabic is that inanimate plurals took feminine singular concord regardless of the gender of the noun:

- (1) Hāḍihi hiya al-kutub al-jadīda.  
 this.FEM COP.FEM DEF-book.PL DEF-new.FEM  
 ‘These are the new books.’

Even in Classical Arabic this semantic overspecification was pragmatically variable, marking a focus on the collective over a more individuated connotation marked with the “expected” plural marking (Brustad 2000, 52–62). It has persisted only variably in the new Arabics. In some it has mutated into related concordial rules, such as in Maltese between the feminine singular and singulative nouns, the latter rendered by secondary affixation of collective-marked nouns.

*Dual marking.* Classical Arabic marked the dual in pronouns, verbal affixes, and adjectival concord. The new Arabics, including Bedouin varieties, usually mark it only on nouns, mostly semantically central ones denoting body parts, time, and quantity. Even here, often the dual-marked noun has become simply plural-marked (the “pseudo-dual”), as in Egyptian *riḡlēn* “legs” rather than “two legs” despite its dual marker (cf. Blanc 1970).

*Genderless numbers.* In Classical Arabic, not only did cardinal numbers 3 through 10 occur in gender-marked pairs, but masculine nouns occurred with feminine numbers and feminine nouns with masculine numbers (numeral *polarity* to the Arabist). This local reversal of gender marking is a complexity common in Semitic. In the new Arabics, the masculine and feminine forms are reallocated in a less arbitrarily complex fashion, as short and long forms, respectively. The long forms are used in isolation (such as for counting) and the short ones as modifiers. In some dialects the devolution has proceeded so far that only one form remains for cardinals, such as in Tunisian, Cypriot Maronite, and some Sudanese dialects (Kaye 1976, 174). Other than the exceptional retention of the polarity in Gulf Arabic (and even here receding today), predictably, it is in a Bedouin variety that Johnstone (1967) reports vestiges of the feature.

*Gender in second-person and third-person plural.* Classical Arabic distinguished gender in the second- and third-person plural in its pronouns and verbal affixes (*?antum* “you pl. masc., *?antunna* “you pl. fem.”; *hum* “they, masc,” *hunna* “they, fem.”). This feature is vanished in most sedentary Arabics, and where it persists is recessive, such as in Sudanese where women retain it but only optionally.

*Analytic possessive.* Classical Arabic placed the burden of the possessive functional load upon the classic Semitic construct: “the king’s book” *kitābu lmaliki*. In the sedentary Arabics, the construct is generally restricted to body parts and family members, while an analytic construction with a particle is used elsewhere, and often more prevalent in the grammar as a whole. Example: Gulf Arabic’s particle *māl*, as in *il-lōn māl wajh il-mara* “the color of the woman’s face” (Holes 1984, 91). Except in Eastern Libyan (according to Owens’s description), the analytic construction dominates in all varieties except Bedouin ones (Rosenhouse 1984, 42, 47), where it is used mostly for clarification especially in NPs with multiple nouns.

*Tense-blind negation.* Negative marking of verbs varied in Classical Arabic according to tense:

present	<i>lā taktubu</i>	‘you do not write’
past	<i>lam taktub</i>	‘you did not write’
future	<i>lan taktuba</i>	‘you will not write’

*Lā* was also used in the negative imperative:

negative imp.	<i>Lā taktub</i>	‘don’t write’
---------------	------------------	---------------

No new Arabic retains this distinction. Instead, variants of *mā* cover all tenses, with *lā* restricted to the negative imperative, most robustly in eastern varieties. Thus in Syrian:

- (2) **Mā** bəlḥa”  
NEG FUT.I.catch up  
‘I won’t have time.’
- (3) **Mā** ḥabbōt(h)a  
NEG loved.I-her  
‘I didn’t love her.’ (Brustad 2000, 284)
- (4) **Lā** tāxid(h)a  
NEG you-take-her  
‘Don’t marry her.’ (ibid., 294)

*Feminine comparative.* In Classical Arabic, the comparative (the Arabist’s *elative* [= comparative and superlative]) of adjectives was formed via nonconcatenative alternations; in the masculine: *kabīr* “big,” *ʔakbar* “bigger.” But the feminine form of the comparative was generated via a quite different pattern: *kubrā*. Already functionally restricted in Classical Arabic, this form has disappeared in all new Arabics. In this, these Arabics lessen the degree of fissional (redundant) morphology, and as such decrease structural elaboration in the grammar via increasing transparency (as defined, for example, by Kusters [2003, 26–30]).

*Analytic comparative.* With some adjectives, new Arabics eschew the nonconcatenative comparative forms in favor of allowing the expression of the comparative via modification of a bare adjective with a form for *more*. For example, Gulf Arabic:

- (5) Jāsim taʕbaan akḥar min ʕAli.  
Jaasim tired more than Ali.  
‘Jaasim is more tired than Ali.’ Holes (1984, 93)

The nonconcatenative comparative pattern was not categorically applied in Classical Arabic either, but in that language, was replaced not with the bare adjective but a nominal form, and specifically the *indefinite* and *accusative* one: *muʔtahid* “diligent,” *ʔakḥar iʕtihādan* “more diligent,” that is, “more diligence.” Furthermore, while participial adjectives tended to occur as infinitives in the comparative, there were unpredictable cases where they occurred as a nominal instead (Schulz, Krahel, and Reuschel 2000, 351). The new Arabic situation, then, is less complex than its antecedent.

*Relative marker.* The Classical Arabic relative marker declined for person and number: *aladhī* (masc. sing.), *alati* (fem. sing.), *alaḏīna* (masc. pl.), *allātī* (fem. pl.) (there were also dual forms). Modern Arabics have invariant morphemes, often variants of *illi*.

*Bimorphemic question words.* Some new Arabics have developed question words based on the conglomeration of forms for *what* with modifiers; the degree of this trait varies across the languages. Moroccan is typical of one Arabic where the process is robust (Caubet 1993, 172):

*āš* ‘what’  
*el-āš* ‘why’ (on what)  
*kif-āš* ‘how’ (like what)  
*wəqt-āš* ‘when’ (time what)

Analogous to similar question words in many pidgins and creoles, these forms undo unitary lexicalizations in favor of the transparency of compositional semantics.

*Loss of internal passive.* In Classical Arabic, the Measures are rendered passive with vocalic patterns—regularly, [u-i] in the past: *kataba* ‘he wrote,’ *kutiba* ‘was written’; and [u-a] in the present: *yaktubu* ‘he writes,’ *yuktabu* ‘is written.’ In all of the modern Arabics but the Bedouin varieties, vocalic indication of the passive is lost except for in small, nonproductive subsets of verbs or in cases due to influence from Modern Standard Arabic (e.g., Kaye and Rosenhouse 1997, 297–98). In Classical Arabic, Measure Seven typically had passive semantics inherently, this expressed with an *-in* prefix (*inqalaba* ‘he/she/it was overturned’). Productively, New Arabics denote the passive with prefixation derived from this Classical Arabic Measure (Maltese *sab* ‘find,’ *instab* ‘was found’ [ibid., 297]) or Eight with its [t]-based prefix, such as Gulf *ḏakkar* ‘to remind,’ *tiḏakkar* ‘to remember.’ Indeed, affixation is not more complex than vocalic alternation in itself—but because the aspect-sensitive differentiation between the Classical Arabic [u-i] and [u-a] patterns is lost, the New Arabics have lost a degree of structural elaboration.

*Loss of subjunctive and jussive.* Classical Arabic marked the subjunctive morphologically, the mood also requiring complementizer *?an*:

- (6) *hal turīdu ?an aṣḥabaka?*  
 you want that I.accompany.you  
 ‘Do you want me to accompany you?’

The jussive was also marked morphologically, used in imperative and hortative constructions as well as with the negated past. Both morphologies are absent in all of the new Arabics.

Moreover, to parse this loss simply as a result of erosion of unstressed final vowels leaves the question as to why the *?an* complementizer was also all but lost (Kaye and Rosenhouse 1997, 307). Analysts are given to supposing that this disappearance was due to the loss of the morphology, such as Kaye and Rosenhouse’s (ibid.)



passing explanation for the loss of *?an*. But this result was hardly inevitable: *?an* could plausibly have remained as a free marker of the subjunctive.

This point also applies to the related case of the tense-sensitive negation. This, too, comprised morphological mood marking: negation in the past and imperative occurred with the jussive, and in the future with the subjunctive (these moods’ markers overlapping considerably, and hence both’s *taktub* in the third-person singular, although the full form of the subjunctive was *taktuba*). However, to suppose that the loss of this marking predicted the loss of the differentiation in negators would seem to contravene a basic principle of language change, in which the loss of a redundant marker often leaves behind one to carry the full functional load of the distinction.

Umlaut marking is prototypical. To assume that the eclipse of jussive marking spelled the disappearance of the distinction between present-tense negator *lā* and past-tense negator *lam* requires that when plural suffix *-i* was lost before Old English, the umlaut it had conditioned in nouns like *gōsi* “gesse” (<*gōsi*) retreated and left the original [ō]. The new Arabics themselves offer similar cases. Classical Arabic distinguished imperfect and perfect verb stems with both vocalic distinctions and affixal ones. Moroccan collapses the vocalic distinctions, but this has not led to the loss of affixal distinction, and, instead, leaves these to carry the semantic load:

TABLE 7.1. Morphological Distinction  
of Imperfect and Perfect in Classical  
and Moroccan

	Classical	Moroccan
I wrote	<i>katab-tu</i>	<i>ktāb-t</i>
you (m.) wrote	<i>katab-ta</i>	<i>ktāb-ti</i>
I write	<i>?a-ktub-u</i>	<i>n-ktāb</i>
you (m.) write	<i>ta-ktub-u</i>	<i>t-ktāb</i>

In this light, we might also expect that at least some new Arabics developed paradigms of tense-sensitive negative prefixes from the free allomorphs in Classical Arabic. But not a single one has. This suggests a simplificatory imperative guiding all of these varieties.

SVO. While word order has been variable at all stages of Arabic according to pragmatic functions, Classical Arabic was largely VSO, while the new Arabics are largely SVO—although the VSO tendency is markedly more robust in the Bedouin varieties. Evidence that subject-first ordering is generationally primary in human speech includes its vast prevalence worldwide and its almost universal occurrence in creoles regardless of the word order of their substrate languages.

*Verb patterns retained.* The only new Arabic that retains all ten of the Classical Arabic verb measures is—surprise!—Bedouin, the Western dialects of which have even developed some new measures (Rosenhouse 1984, 32). Otherwise, the loss of the

Fourth Measure (passive/causative) is regular, and in many dialects, various other measures are only weakly productive. Chadian Arabic retains only three forms productively, I, II, and VII (Kaye and Rosenhouse 1997, 293). Thus the new Arabics tend to have shed a significant degree of allomorphy and fissional morphology.

There are further reductions whose particulars vary according to dialect. For example, in Classical Arabic, when *w* is the first consonant in a root, only in the perfect does the verb inflect regularly. In the imperfect, it drops out: perfect *waṣal* “he arrived,” imperfect *yaṣil*, where *ya-* is a prefix and the *w* is absent. When *w* is the second or third consonant, in both the perfect and imperfect it tends to both drop out and lengthen the proceeding vowel (from root ZWR, *zāra* “he visited”). Moroccan irons out this wrinkle considerably: when *w* (or other weak consonants *y* or *ʔ*) is first, the verb remains regular; when it is second, the verb is regular in the imperfect but not in the perfect; only when it is third is the verb irregular in both perfect and imperfect (cf. Kusters 2003, 144). Also, Classical Arabic verbs fell into two main vowel patterns in the perfect, *a-a* (*kataba* “he wrote”) and *a-i* (*šariba* “he drank”). Moroccan, as noted above, has eliminated these patterns: *ktāb*, *šrāb* (e.g., Kaye and Rosenhouse 1997, 294).

### 3.5. *Implications of the Table*

It is tempting to suppose that the losses in the above table are natural to how language changes over time. Arabists typically have supposed so, such as Cohen’s (1970, 111) assuming that such developments “can in reality be considered as a common line of evolution, a tendency of the language.”

But this perspective is valid more for tendency than inevitability. When we pull back the camera, so to speak, from Romance and Germanic, the descendants of Arabic appear more anomalous.

Bantu speakers are assumed to have begun their migration into central and southern Africa in 2000 BC with a proto-language containing an array of almost two dozen noun classes. Four millennia later, hundreds of Bantu languages survive with most of them intact. The Indo-European family is reconstructed as being between ten and eight millennia old, and today most of its descendants retain enough of its morphological elaboration that the ancestral language can be confidently reconstructed in all of its inflectional might. The analytic tendency of Romance and much of modern Germanic is exceptional: highly inflected descendants like Baltic, Slavic, and Greek preserve their inflectional legacy robustly, and Celtic and Indo-Aryan do so to a lesser but hardly vestigial degree. More locally, in Slavic only one language, Bulgarian, has lost its nominal affixes, while all of the others retain this plus the ample verbal morphology of the family that even Bulgarian retains.

Why, then, have *all* of the descendants of Arabic tossed off all case marking and all morphological marking of mood in just a millennium and a half?

Crucially, our question is not why just *one* variety might take this path: things happen. Our question is why such a degree of reduction would occur so uniformly. It would seem that under “normal” conditions, perhaps Moroccan Arabic would be as it is, but then over in Egypt and up in Syria, at least some case marking would

persist, while Sudanese and Gulf would maintain some of Classical Arabic's subjunctive marking; a variety or two would continue to decline the relative marker while some other ones would still be VSO; at least a few sedentary varieties would hold to the synthetic "construct" possessive, and so on. But this is not the case.

Why could not a single new Arabic continue to harbor, for example, negation markers linked to tense, given that exactly while the new Arabics were emerging, in China Cantonese and Min were multiplying the battery of negative markers they inherited from Old Chinese, linking them to ever-finer semantic distinctions?

## 4. Innovations

### 4.1. *Innovations in the Sedentary Arabics*

Given that an episode of interruption of transmission does not stop a natural language from going on afterward to mark new categories and develop further complexities, we would not expect that the new Arabics would remain subtractions of Classical Arabic in their totality.

In the example most often cited, the new Arabics have conventionalized particles, used usually with imperfective verbs, to connote various tense and aspect distinctions more explicitly than was necessary in Classical Arabic (although it contained precursor usages of such particles), such as Gulf examples *b-adris* "I will study," *gʕid yilʕab* "he is playing" (from the "sit" verb), *kān yišta yil* "he used to work." The new Arabics thus leave such distinctions to context less than their ancestor did, and are thus more overspecified in this regard.

Also, the tendency for many new Arabics to restrict the construct in the possessive to body parts and family members while using the analytic possessive elsewhere, while under one perspective demonstrating a leaning toward analyticity typical of non-native rendition, also lends these dialects an alienable/inalienable distinction. Analytic possessive marking was not unknown in Classical Arabic, but was much less widespread than in most of the new Arabics and served contrastive rather than semantic function, as shown in this Modern Standard Arabic example (Brustad 2000, 71):

- (7) (a) markaz al-luḡāt                      al-yamanī  
           center DEF-languages DEF-Yemeni  
           'The Yemeni Language Center' (not Egyptian)
- (b) al-markaz al-yamanī li-l-luḡāt  
           center DEF-Yemeni for-DEF-languages  
           'The Yemeni Language Center' (not Cultural)

Many of the new Arabics have replaced the indefinite marking function of the lost case markers with new free morphemes, such as Iraqi's *fard* or ones derived from "one" such as Moroccan's *wāḥad l-*. These are, however, still at a less grammatical than pragmatic stage in their evolution, translating roughly as "a certain" or "some" and used for vivid individuation within a narrative in the vein of topic

markers. Brustad (2000, 18–43) classifies these markers as intermediate between definiteness and indefiniteness as traditionally defined, occupying only part of the domain of the unspecified that defines indefinite marking per se. For example, Syrian and Egyptian's versions are only used with humans, and, in general, abstract and generic nouns are marked as definite despite being unspecific, such as at the opening of this Syrian joke where restaurant is marked definite despite its indefinite status in an English sentence where the indefinite article is more deeply grammaticalized (Brustad 2000, 25):

- (8) Fī wāḥid badwi, fāt ʕa l-maṭṭam.  
 there-is one bedouin entered.he into DEF-restaurant  
 'There's this bedouin who went into a restaurant.'

Elsewhere, the new Arabics have developed new structural elaboration, in the morphosyntactic sense. In Classical Arabic, most feminine nouns' singular marker was underlying [at], but expressed as [ah] before a pause (including in citation form). Thus "pack" is *ʕulbah* in citation form, but elsewhere carries the [t] (the *tā' marbūṭa*):

- (9) ʔaʕṭīni ʕulbat-a saǰāyir.  
 give.IMP pack-ACC cigarette.PL  
 'Give me a pack of cigarettes.'

In the new Arabics, the [t] has become the more environmentally particular "hidden *t*," appearing only in certain bound environments. For example, in Gulf Arabic it appears in constructs (*nifta-t xubuz* "a bit of bread"), before suffixed pronominals (*sāʕa-t-i* "my watch"), and on numerals with vowel-initial nouns (*sitta-t awlād* "six boys"). Thus a simple interaction between syntax and phonology in Classical Arabic has devolved into several morphophonological ones in the new Arabics, which can be seen as a complexification of their grammars.

In Classical Arabic, verbs in the perfect in Measure One occur in two main vowel patterns, *a-a* (*kataba* "he wrote") and *a-i* (*šariba* "he drank") (an *a-u* type is marginal). While there are cases like Moroccan that collapse these into one (*ktāb, šrāb*), Maltese multiplies them into six (*talab* "he asked," *qatel* "he killed," *fehem* "he understood," *seraq* "he stole," *yibed* "he pulled," *ħolom* "he dreamt") (Kaye and Rosenhouse 1997, 294).

As is well known, simplification often leads to new complexities, the emergence of phonemic umlaut after inflectional erosion being a classic example. While there are no logical grounds for assuming that simplification must necessarily generate an equivalent degree of "compensatory" complexification, predictably in the new Arabics, simplifications sometimes create new complexity, such as in Moroccan, where one symptom of the erosion of short vowels is that CV syllables are disallowed with short vowels. In many paradigms, this leads to metatheses that keep short vowels sequestered behind the consonant codas that render them legal. Notice below how in the first and second persons, [ɔ] is contained within a syllable with consonant coda in the first and second persons, but can only be thusly contained in the third person via a metathesis that avoids the illegal \**ktābu*. The result is an allomorphy absent in the Classical ancestor (Versteegh 1997, 166):

TABLE 7.2. New Allomorphy in Moroccan Verb Morphology Compared to the Classical Source

	Classical	Moroccan
we wrote	<i>katabnā</i>	<i>ktābna</i>
you (pl.) wrote	<i>katabtum</i>	<i>ktābtu</i>
they (m.) wrote	<i>katabū</i>	<i>kātbu</i>

4.2. *Sedentary Innovations in Perspective*

If the sedentary Arabics had not complexified at all during their mutation over a millennium and a half, then we would be faced with a distinctly perplexing stasis unknown to any natural languages. Yet it must be acknowledged that the above developments hardly bring these dialects to par with their ancestor in complexity.

The new Arabics remain abbreviations of their phonological and morphological heritage. In the latter case, they tend strongly to eliminate categories rather than replace them with analytic equivalents: the construct is replaced with the analytic possessive, but the dual, the plural second- and third-person markers, and the jussive are left unexpressed. The particular semantic space that the latter marking carved out, occupied by the negative imperative and the hortative, is swallowed up by the more general imperative in the new Arabics. The encroachment of analyticity upon the morphological comparative takes its cue from a parallel phenomenon in Classical Arabic, but unravels the latter's sensitivity to case, definiteness, and more in favor of creole-style transparency. The linkage of negative marking to tense is extinct, and verb morphology remains in all cases less elaborated than in Classical Arabic.

If the concurrent innovations in these dialects indicated nothing unusual in their timelines, then we would expect that the Arabics learned least by foreigners, the Bedouin dialects, would display like degrees of innovation, but not more. But this is not the case. It is precisely in the Bedouin dialects that we see not only more retention of ancestral features but also the elaborifications upon these that are familiar from cases like Slavic, with its entrenched distinction of perfective and imperfective in verb stems. To wit, the Bedouin Arabics have more the profile of languages "left alone" to mutate unimpeded than varieties like Moroccan and Egyptian.

For example, Bedouin varieties, while losing the Classical Arabic subjunctive and jussive markers, have concurrently developed a battery of preverbal modal morphemes that lend them a more elaborated expression of the pertinent nuances than their ancestor had (Ingham 1994, 119–29). As noted above, Western Bedouin dialects not only retain all ten of Classical Arabic's verb measures but also have even developed some new ones. The Yemenite variety has four vowel patterns in perfect verbs: *ragad* "he slept," *širib* "he drank," *wugif* "he stopped," and *ʔuṭus* "he sneezed" (Kaye and Rosenhouse 1997, 294). Concurrent with these developments, Bedouin varieties share sedentary elaborifications such as preverbal particles and a morphophonemically conditioned remnant of the feminine final *t*.

### 4.3. *New Semitics Elsewhere*

The anomaly of the sedentary new Arabics is most usefully illustrated through comparison with the changes that other Semitic languages have undergone in mutating beyond their ancestral stages. A claim that the contrast between Classical and Moroccan Arabic is ordinary is challenged by, for example, Modern Aramaic dialects, whose divergences from Old and Middle Aramaic in no sense parallel those from Latin to Spanish or Proto-Germanic to Dutch.

#### 4.3.1. ARAMAIC

The modern Aramaic dialects are isolated varieties spoken in small communities, and thus represent developments of Semitic little interrupted by non-native acquisition. In confirmation of the argument of this monograph, they represent a mutation of, rather than an overall abbreviation of, their ancestor, in the same way that Slavic does of Proto-Indo-European. The guiding theme of descriptions of the pathway from Old to Modern Aramaic is transformation rather than loss. There is hardly retention, and the losses are amply counterbalanced by innovations that mark new categories and complexify the morphosyntax. This is demonstrated below with data from the Ma'lūla dialect of Western Aramaic (Jastrow 1997).

#### *Retentions*

1. Western Aramaic retains a highly conservative phonemic inventory, closely paralleling that of Middle Aramaic and retained over the millennium since the transition from that stage (500–1000 AD). This includes five short vowels surviving robustly, in contrast to their weakness in the new Arabics.
2. Gender distinctions in the plural of verbal paradigms are retained.
3. The Classical Aramaic count plural is retained in use with numbers, complete with gender distinctions:

	<i>door</i> (m.)	<i>paper</i> (f.)
singular	<i>θarʕa</i>	<i>war ʔkθa</i>
plural	<i>θarʕō</i>	<i>warḵō,a</i>
count plural	<i>θarʕ</i>	<i>warḵan</i>

4. A Classical Aramaic distinction in nominal morphology between absolute and emphatic survives as indefinite and definite marking on adjectives, varying according to number and gender, as with the word for *poor*:

	indefinite	definite
masc. sing.	<i>ifker</i>	<i>fķīra</i>
fem. sing.	<i>fķīra</i>	<i>fķīrča</i>
masc. pl.	<i>fķīrin</i>	<i>fķīro</i>
fem. Pl.	<i>fķīran</i>	<i>fķīrθa</i>

5. Gender distinction is preserved in numerals up to nineteen.
6. There are nine productive verbal measures.

### *Innovations*

1. Middle Aramaic stops had fricative allophones that phonemicized in the modern varieties, yielding a richer consonantal inventory that includes [f], [č], [θ], [ð], [x], and [ɣ].
2. Umlaut, conditioned by lost final high vowels that raised preceding mid vowels, lends new allomorphy to verb paradigms. For example, an extinct feminine imperative suffix *-i* has left behind an umlaut distinction: “open!” *fθōḥ* (m.), *fθūḥ* (f.).
3. While retaining the Semitic perfect and present as the “past” and “subjunctive,” modern Aramaic has innovated new present and perfect paradigms by remodeling active and passive participles respectively.
4. In the new present and perfect, there is a gender distinction not only in the second- and third-person plural but also in the first-person plural, unusual in Semitic.
5. Typical of affixes that have been in uninterrupted use by native speakers over long periods of time, Aramaic suffixes often condition more phonetic disruption than new Arabic ones. The richness of the morphophonemics is demonstrated by cases like the third-person singular object suffix *-e*, which conditions the elision of the root’s first vowel and metathesis of the second consonant and second vowel: *ifθaḥ* “he opened,” *faṭḥe* “he opened it.” In the new present and perfect, object suffixes occur after an intermediary [l], and shift stress forward while conditioning vowel changes: *nōṣka* “she kisses,” *naṣkōle* “she kisses him.” Historical layering further enriches the morphophonemics, as suffixes often occur with earlier forms of the root no longer current otherwise: *faṭṭḥēṣunne* “you (m. pl.) opened it” is based on the earlier *\*faṭṭḥēṣun* rather than modern *faṭḥiṣun*.
6. Definite objects are marked with a verbal suffix: *ṣatṭril yaḇrōna* “he sent the man.”

There are similar phenomena in Eastern Aramaic dialects, such as ergative inflection of the verbal base and paradigms of both free and enclitic copulas inflected for present, past, and negative (Jastrow 1997, 373–74).

### 4.3.2. SOUTH ARABIAN

The contrast is similar between new Arabics and the Semitic languages of Arabia that have remained isolated while their relative Arabic spread vastly and underwent extensive non-native acquisition. The six South Arabian varieties—distinct languages rather than dialects of a single one—are spoken in Yemen and Oman mostly by seminomadic peoples, often living in mountainous areas. There are certainly losses: there is no feminine concord with generic plurals or tense-sensitive negation, and the construct is largely frozen and restricted to inalienables. Yet like Modern Aramaic, these languages overall demonstrate “normal” outcomes of language change, with rich retention and innovation.

They have, for example, the most conservative consonantal inventory in Semitic, including ejectives, three alveolar fricatives, and, in some, retroflex articulation when

*r* is followed by denti- or lateral-alveolar consonants. There is dual marking in all three persons, on nouns, pronouns, and in verb paradigms (although the robustness of this marking is variable in many of the languages). Gender marking persists in the second- and third-person plural. There are ten verb measures including a passive one, and subjunctive marking is retained.

The languages innovate conditional marking, and while the construct is moribund, a new particle binds determinatum to determiner. As in Aramaic, suffixed pronominals wreak especial havoc on pattern, timbre, quantity, syllabic structure and stress: “he beats” *isūbat*, “he beats her” *isūbtas*; “nose” *naxrīr*, “my nose” *anxráyri* (data from Simenone-Semelle 1997).

## 5. Implications for Related Analyses in the Arabist Literature

### 5.1. Case and Definite Marking (and Beyond)

The 1970s witnessed arguments that the loss of the case and definite suffixes in the new Arabics signaled nothing extraordinary in terms of language change, and had in fact already occurred before the spread of Islam. Corriente (1971) traced the disappearance to their semantic redundancy, while Zwettler (1978) characterized the loss as a symptom of widespread erosion of final short vowels. Arguments such as these can be taken as representing a strain of thought seeking to “demystify” the changes from Classical to New Arabic which appear, on first glance, decidedly thorough. Representative is Cohen’s abovementioned comment (addressing the changes Ferguson [1959] adduced arguing for a common koine ancestor, on which see below) that “all these phenomena can in reality be considered as a common line of evolution, a tendency of the language.”

#### 5.1.1. ARGUMENTS AGAINST THE LOSSES AS “NATURAL”

But these claims about the suffixes in question did not stand up to persuasive linguistic and documentational objections. Short vowel loss cannot explain why the oblique plural case/definite suffixes were also eliminated, since they had long vowels (*mudarrisīna* “teachers” [acc./gen.]) (Versteegh 1984, 4). All languages are rife with redundancy, quite often of long standing (*ibid.*, 91) (many Indo-European languages harbor redundant concord marking several millennia old). Versteegh also objects that if the markers were already lost before Islam, then we would not expect anecdotes in documents mocking foreigners’ lack of control of these very markers (11). Relatedly, Joshua Blau (1977, 15) noted that if new Arabic had already arisen when the Qur’ān was composed, then we would expect at least some interference from such varieties in that document as there tends to be in later Classical Arabic writings, and yet there is none.

The new Arabics’ loss of subjunctive marking is similarly inapplicable to phonetic erosion, as Versteegh (1984, 90) notes. Many of the subjunctive endings were indeed short vowels ripe for elimination. But then many were long vowels, and in others, the short vowel was preceded by an *n*. Yet the new Arabics swept away the



endings in all persons and numbers. A phonetic account leaves unexplained why there could not have arisen a paradigm where long vowels persisted and the *-na* suffixes, for example, when shorn of *-a* took on an anaptyctic schwa before the *n*:

Table 7.3. Subjunctive Marking in Classical and a Hypothetical New Arabic

	Classical	Hypothetical New
I write	<i>ʔaktuba</i>	<i>ʔaktub</i>
you (m.) write	<i>taktuba</i>	<i>taktub</i>
you (f.) write	<i>taktubī</i>	<i>taktubī</i>
he writes	<i>yaktuba</i>	<i>yaktub</i>
she writes	<i>taktuba</i>	<i>taktub</i>
we write	<i>naktuba</i>	<i>naktub</i>
you (m. pl.) write	<i>taktubū</i>	<i>taktubū</i>
you (f. pl.) write	<i>taktubna</i>	<i>taktubən</i>
they (m.) write	<i>yaktubū</i>	<i>yaktubū</i>
they (f.) write	<i>yaktubna</i>	<i>yaktubən</i>

That phonetic erosion need not overapply, and can instead leave partial paradigms, is indicated by spoken French, in which only first- and second-person plural marking remains in two of its verb classes in the present tense (*parlons*, *parlez*), and in English, in which only the third-person singular is marked in the present tense. Indeed there is no better demonstration of the plausibility of the above hypothetical paradigm in that Classical Arabic itself had one quite similar in the jussive:

I write	<i>ʔaktub</i>
you (m.) write	<i>taktub</i>
you (f.) write	<i>taktubī</i>
he writes	<i>yaktub</i>
she writes	<i>taktub</i>
we write	<i>naktub</i>
you (m. pl.) write	<i>taktubū</i>
you (f. pl.) write	<i>taktubna</i>
they (m.) write	<i>yaktubū</i>
they (f.) write	<i>yaktubna</i>

5.1.2. THE LOSSES AS SUPPORT FOR NON-NATIVE ACQUISITION

The refutation of the Coriente/Zwettler interpretation positions us to view the loss of the case/definite markers more broadly (as well as by extension other markers such as the modal ones above). If the case/definite markers were not already lost before Islam, then this implies that they were lost in the various new Arabics across a vast expanse of territory. As Blau, Versteegh, and others have noted, this loss cannot be seen as inevitable, and this leads to the question as to why the loss occurred in every

single new Arabic, with even Bedouin varieties harboring little more than a remnant, as opposed to retention, of the original system.

Indeed, Proto-Semitic case marking has been rather evanescent over the millennia: the only members retaining a three-way distinction are one now extinct (Akkadian) and another artificially frozen at an archaic state (Classical, and thus Modern Standard, Arabic); Ge'ez retained an accusative marker in the singular and East Gurage today restricts this further to the definite (Hetzron 1987, 659). Surely, if we conceptualize the trait as one that could survive in a mere two descendants of the protolanguage, it seems unremarkable that no modern descendant retains it today.

But this is perhaps a distortive view. Semitic has never been a very large family, constituting about three dozen languages today depending on one's conception of language versus dialect among Gurage and new Arabic varieties. Given that Semitic is generally assumed to have emerged roughly five thousand years ago in about 3000 BC, presumably Semitic had yielded even fewer languages two thousand years ago, when Arabic had yet to split into new languages and Ethio-Semitic is assumed to have been at its founding. In this light, the persistence of case marking in two members of the family in its new and compact state at that time indicates less fragility than it would in, for example, vastly multiple language groups like Bantu or Oceanic. And with this acknowledged, we must ask why the markers have been utterly lost in every single sedentary descendant of Classical Arabic, maintained fully, or even approximately, in no modern variety at all including the Bedouin ones. If three thousand years into Semitic's existence, two members of the still smallish family maintained Proto-Semitic's full nominal declension system, then we would expect that after another two thousand years, among the descendants of one of those two languages, the number retaining the full system would be, at least, more than none. This is so, especially given how many such new Arabics exist, such that there have been dozens of "rolls of the dice." Certainly we would not expect the utter absence of the system in any sedentary variety at all.

That is, unless something besides ordinary transmission created these new Arabics. It is here that Versteegh's (1984) approach, treating the disappearance of these markers as a symptom of non-native acquisition, becomes all but conclusive.

## 5.2. *The Koine Question*

### 5.2.1. MIGRATION AS A CHECK ON DIVERGENCE?

Diem (1978) addresses the question as to why the new Arabics have diverged so little from the Classical Arabic template, compared to Old Aramaic's split into distinct and highly innovative varieties and the profound remodelings of Proto-Semitic equipment in the Ethio-Semitic varieties (*ibid.*, 130). He identifies the nomadic lifestyle of early Arabic speakers as the reason for the uniformity, supposing that they developed a common koine for communication purposes, their travels preserving this over time, pointing to the relative uniformity of Bedouin dialects today as support for the argument (*ibid.*, 132). Joshua Blau (1965) made a similar point.

Diem is correct in rejecting arguments that prescriptive influence from the Qur'an retarded normal change in the new Arabics (e.g., Cohen 1970, 124). Under

this analysis Jewish varieties of Arabic and rural ones, less influenced by literacy, would be further from Classical Arabic when in fact they are not (Diem 1978, 139–43). However, it is less plausible that contact can usefully explain the relative closeness of all of the new Arabics to their ancestor.

Diem’s observations about certain population movements within the Arabic-speaking territory are well taken. Yet there is a question as to how plausibly these can have decisively affected the casual speech of every single Arabophone across the vast swatch of territory from Morocco to the Middle East. It is doubtful that widely separated Bedouins from Algeria to Arabia have at any point migrated in such large numbers, with such frequency, and so uniformly across the relevant territories to homogenize casual speech among untold millions to such a striking degree. The sum of the movements would presumably have led to a certain degree of convergence, but accompanied by pockets of distinctly divergent Arabics where the migrations happened not to reach. Where are the isolated Arabs whose everyday language is a mutation rather than a reduction of Classical Arabic, analogous to the villagers today using an Aramaic just as complexified as the language was a thousand years ago?

It is also relevant that the new Arabics are in certain fashions oddly *dissimilar* as presumed products of a drive toward intercomprehension. Since to the untutored speaker, words are the most prominent aspects of a language, the first thing we would expect of varieties born amid a leveling impulse would be at least some evidence of a drift toward lexical similarity. Yet this is precisely where the new Arabics diverge most, as we see from the word for *now* in various new Arabics (Kaye and Rosenhouse 1997, 309):

Table 7.4. *Now* in New Arabics

Nigerian	<i>hatta</i>
Egyptian	<i>dilwaʔti</i>
Syrian	<i>hallaʔ</i>
Iraqi	<i>hassa</i>
Gulf	<i>dahḥīn</i>
Maltese	<i>issa</i>
Moroccan	<i>daba</i>
Algerian	<i>drūk</i>
Tunisian	<i>tawwa</i>
Libyan	<i>alān</i>
Israeli Bedouin	<i>ʔessa</i>
Mauritanian Bedouin	<i>dark</i>

Indeed, the non-Arabist such as myself must accustom himself to the general sense in the literature that the list above represents varieties that are considered, in any sense, “dialects” of one another, with endless observations that the speech of, say, a Moroccan is “barely comprehensible” to the Iraqi, as if we were dealing with highly divergent renditions of a single “language.” Rather, as Chinese speakers harbor a sense that their various languages are dialects of one another because they share a writing system, I often sense that the shared heritages of Islam, the Modern Standard language, and the writing system distract Arabic specialists from openly address-

ing the new Arabics as the distinct *languages* that they are, every bit as distinct as the Romance languages and much more so than the Scandinavian ones. In this light, the new Arabics no more suggest koineization than do French, Spanish, and Italian, despite the very similar grammars that they share. After all, Spanish is fitfully but “barely” comprehensible to the Italian—yet we do not suppose that the grammatical similarity between Spanish and Italian was due to efforts at mutual comprehension between their speakers amid medieval population movements in Europe.<sup>4</sup>

### 5.2.2. ABBREVIATION AS A CHECK ON DIVERGENCE

I would venture that the new Arabics are as little diverged from their ancestor as they are because they are all the products of abbreviation of their ancestor’s structure, followed by a relatively brief period in which they have had the time to develop new complexities.

Structurally, languages differ most in the idiosyncratic modules of phonetics, phonology, morphology, and morphophonemic processes (such that these, for example, present the greatest challenge to adult learners). In syntax, variation is more confined, such as to permutations of word order among the three components of subject, verb, and object, and alignment choices such as nominative/accusative versus ergative/absolutive. The new Arabics were born amid vast abbreviation of Classical Arabic’s phonetic, phonological, and morphological distinctions. This left all of them with less material susceptible to idiosyncratic elaboration than Aramaic dialects, carrying the full battery of their previous stages’ folderol, such as fricative allophones ripe for phonemicization, or an absolute/emphatic distinction in nominal morphology reinterpretable as an affixal indefinite/definite distinction in adjectives. All Arabics have moved toward complexity over the past millennium and a half, but this has not been long enough for them to equal their ancestor in the degree of complexity.

One might surmise that at a future stage, the new Arabics might have paradigms of prefixes marking tense, mood, and aspect evolved from free morphemes marking these categories today. The analytic possessive morphemes could similarly prefix to their referent, while an alienability distinction could become regularized, the result being grammaticalized alienable possessive markers. Truly grammaticalized indefinite markers could emerge (and perhaps then become affixes like the definite marker). Phonologically conditioned allomorphies could muddy paradigms to a degree surpassing exceptionality and moreover drift into morphologization, as in Slavic.

But this will require time, and we encounter the new Arabics at a stage when they have developed but a moderate degree of innovations upon these most idiosyncratic aspects of a grammar. The new Aramaics, having never experienced an interruption in transmission, represent the mutation of Old Aramaic in the totality of its millennia of accretions not only syntactic but also phonological and morphological, such that earlier idiosyncrasies have evolved into new ones. The new Arabics represent what happens when a human language has been shorn of a degree of its idiosyncrasy. Naturally, being in the process of restoring that which most starkly distinguishes grammars from one another as well as from their ancestors, the new Arabics resemble their ancestor more than other languages of a similar age.

Importantly, this explanation accounts for the extreme lexical divergence among the varieties. Under this account, intercomprehension was not a significant factor in their development. Their structural similarity is due to predictable effects of birth amid non-native acquisition, within which structural idiosyncrasy suffers. Meanwhile, their lexicons have diverged in the fashion expected of sister languages evolving separately.

### 5.2.3. FERGUSON'S KOINE

While I find Diem's postulation of a koine as an explanation for why the new Arabics are so similar unconvincing, koineization is not completely irrelevant to these varieties' history. Specifically, Ferguson's (1959) tracing of the sedentary Arabics to an initial rather than gradually developing koine is a more likely scenario.

Actually, the new Arabics would be equally reduced if no such common koine had existed—non-native acquisition in separate locations could easily have conditioned like abbreviations. However, contrary to Cohen's (1970) claim that the fourteen traits Ferguson adduces are unremarkable results of "drift," many of these traits are idiosyncratic to a degree that suggests common ancestry, as Kaye (1976, 142–63) cogently argues.

By my count, six of Ferguson's traits could plausibly have occurred in several varieties separately. For example, in almost all sedentary Arabics, the original situation where feminine-marked numerals were used with masculine nouns and masculine numerals with feminine nouns is restructured such that the original masculine numerals ("short" forms) are used when a counted noun follows, while the original feminine ones ("long" forms) are used in isolation. Bloch (1971) notes that this is a predictable effect of a cross-linguistic Wackernagel's Law-style tendency, for monosyllables (in the Arabics, the erstwhile masculine short forms) to be attached to surrounding constituents. The replacement of *šāf* with *raʔā* for "to see" is another example. Since Ferguson wrote, Cowan (1966) and Kaye (1986) have shown that *raʔā* persists in Maltese, Cypriot Maronite Arabic, and various Jewish varieties, such that the prevalence of *šāf* elsewhere could well be attributed partly to chance and perhaps partly to diffusion.

But the other traits do not submit to this kind of treatment, and suggest languages descended from a common ancestor in which particular choices happened to have been made. Why do the sedentary Arabics so consistently happen to velarize the "hidden *t*" that appears in the numbers thirteen to nineteen between the basic numeral and the word for *ten*, as in Gulf's *fourteen*, *arba ʕ-ṭ-a ʕšar* "four-ten"? It is understandable that a variety might shed the feminine comparative form, but why does every single new Arabic do so, despite the feminine remaining a robust category in the language? Similarly mysterious is why the relative particle does not mark gender or number in a single one of the varieties. In all of the new Arabics, verbs whose second and third consonants are identical pattern exceptionally in the first- and second-person perfect, paralleling weak verbs with *y* as the third consonant. These "geminate" verbs are susceptible to irregularities across Semitic, and so the collapse itself is hardly surprising. But a comparative Semitic perspective reveals it as but one of many possibilities, having occurred nowhere but in the new Arabics (Kaye 1976).

A few traits seem especially indicative. Across the new varieties, a Classical Arabic plural adjective pattern with *i* as the first vowel (Classical *kaḇīr*, *kibār* “big”) now usually has no vowel (Syrian *kbār*)—but with surrounding consonantal alterations often corresponding to the loss of an original *u* (some varieties retain this *u*). As Ferguson (1959) argues, this suggests a common ancestor in which the *i* had happened to become a *u*, rather than, for example, drop out, persist as an *i*, or perhaps become an *a*. New Arabics have all innovated a verb *jāb* “to bring” via a particular fusion of a verb “come” *jāʕa* and *bi-* “with”; also, *jāʕa* and another “come” verb *ʕatā* are always extinct rather than surviving.

Finally, the retreat of dual marking to nominal marking only, with its elimination in pronouns, adjectives, and verbal morphology, is more indicative of common ancestry than generally assumed. It is true that dual marking has often eroded across Semitic, is marginal in Akkadian and Hebrew, and is extinct in Ethiopic. This, however, does not explain why such extreme erosion has occurred in *all* new Arabics, and especially in the exact same fashion. To assume that this situation is unexceptional leads to a question as to why, if dual marking is so inherently evanescent, it arises in so many languages in the first place. This question becomes even more urgent in that dual marking is so widespread and solidly entrenched in other language families. If historical linguistics had happened to emerge among Australian Aborigines, there would be no sense of dual marking as fragile on a continent where it is found in roughly nine out of ten languages (Dixon 2001, 68). We might expect that some new Arabics would have lost all dual marking, others would retain it only on nouns, others would retain it here plus in, for example, pronouns and/or verbs, while at least one or two would retain the Classical Arabic equipment in its entirety. It would appear that a single ancestor to the new Arabics restricted its dual marking to nouns and passed on this configuration to its descendants.

Arabists such as Joshua Blau (1977) have objected that some of Ferguson’s features are not quite categorical across the dialects. For example, Gulf Arabic numerals retain the reverse gender assignment, and Joshua Blau (1983) notes that Yemenite Judeo-Arabic texts of the seventeenth century retain the *-un* case/definite marker (although this could well have been a grapholectal trait absent in the spoken language). But these predictable hiccups in the data are not numerous enough to belie Ferguson’s basic argument. Nor can Blau’s argument that the commonalities can be explained by interdialectal diffusion stand: as noted above, it is difficult to see how varieties spoken across such a vast region could be homogenized significantly by scattered population movements of a relative few. Whether the koine developed in military camps in Iraq as Ferguson (1959) supposes remains to be ascertained by future research. But the import of the fourteen features remains: both the Bedouin and the sedentary Arabics are too similar in many respects for such an initial parent language not to have existed.

## 6. What Happened to Arabic?

The data suggest that the spread of Arabic with Islam had a decisively reductive effect upon its structure in the locations to which it spread.

### 6.1. *The Emergence of New Arabics*

The Arabs established military garrisons in the regions they conquered, accompanied by massive migrations. For example, between two hundred thousand and four hundred thousand Arabs mixed with a population of about four million Syrians (Poliak 1938). Eighty thousand Arabs joined eight million people in Egypt, followed by another million Bedouin Arabs migrating there and to other North African regions from 1050 to 1350.

The migrant Arabs encountered speakers of Aramaic, South Arabian, Coptic, Berber, Greek, and Persian. The main locus of contact was in cities (Versteegh 1984, 66), which attracted locals from surrounding areas who needed to learn Arabic to participate in commerce and local culture. Significantly, the first solid documentary evidence of new Arabic is from South Palestinian texts copied in cities in the 700s, with less conclusive evidence from the century before this (Joshua Blau 1977, 16–18). Language contact would have intensified as Arabs married local women (as Vikings did in England). That the new Arabics were subject to widespread non-native acquisition is indicated by many accounts in contemporary literature criticizing foreigners' mistakes in the language (Versteegh 1997, 102). Of course, this does not necessarily document these "mistakes" being passed on to succeeding generations, but it does indicate that non-native Arabic became widespread in these new societies. In the emergence of new Arabics, this would be an obvious first step.

Viewing a specific context, we see Morocco receiving large numbers of the 150,000 Arabic-speaking migrants to North Africa soon after the Islamic Empire spread to the region in 650 AD. Berber nomads relocated to the settlements, in numbers large enough that non-Arabs dominated their population. After the fall of the empire, Berber dynasties ruled from the end of the tenth century into the fifteenth, while abovementioned Arab Bedouins began arriving in large numbers for most of this period, as Arabic also spread to the countryside. During and subsequent to this era, Andalusian Muslims and Jews from Spain also added to the mix, along with 150,000 sub-Saharan slaves taken into the army in the 1500s (historical data from Abun-Nasr 1987; and Versteegh 1997).

The new Arabics naturally incorporated some features of the languages with which they came into contact. The especially pronounced loss of short vowels and reduction of long ones in western varieties like Moroccan, for instance, is surely due to influence from the notoriously vowel-shy Berber languages (Kaye and Rosenhouse 1997, 265). Syrian's third-person plural pronoun *hinne* is likely modeled upon Aramaic's *hinn* (Arnold and Behnstedt 1993). The lack of *wh*-word fronting in Egyptian is traceable to the same feature in Coptic (Versteegh 1997, 106–7), and so on.

Yet generally, substratal transfer into the new Arabics is extremely slight (Diem 1979; Versteegh 1984, 25). This, however, does not speak against the languages' birth amid non-native acquisition, nor does it necessitate surmises such as that Islamic cultural identification would have encouraged Arabs to resist foreign incursions into their grammar (Diem 1978, 134). When a language undergoes heavy non-native, but relatively successful, acquisition, substratal transfer tends to be rather low, in contrast to its prominence when acquisition is at the pidgin or creole level. Réunionnais French, a "semicreole," was created by Malagasies and later adopted by Bantu speak-

ers, and yet has only scattered evidence from these people's languages. Afrikaans contains similarly moderate evidence of the Khoi-San languages spoken by the peoples considered responsible for its reduction of Dutch grammar. Also germane is the light degree of influence from Altaic languages upon Mandarin, despite hundreds of thousands of Altaic speakers having acquired and transformed it.

The Bedouins, on the other hand, were isolated from the core of commercial and cultural activity when the capital of the empire moved from Medina to Damascus in 661 AD, and since then they have had mostly commercial relations with sedentary Arabs. The primary social unit has been the tribe, with intertribal warfare once common, and it has been vanishingly rare for sedentary Arabic speakers to join Bedouin tribes (Ingham 1994 Rosenhouse 1984). These are the kinds of social conditions that typically encourage retention of original features and internally generated innovation.

### 6.2. *A Cline of Reduction*

The concept of an intermediate degree of reduction is often difficult to convey in the linguistics literature. Many responses to Versteegh (1984) object that the new Arabics are not as reduced in comparison to Classical Arabic as pidgins and creoles are in comparison to their lexifiers (e.g., Owens 1997, 126), despite Versteegh's clear explanation that he is treating a light degree of restructuring. In order to be maximally clear in my conception of the place of the new Arabics in a language contact typology, I will use Nubi Creole Arabic as a useful contrastive extreme.

Nubi Arabic is indisputably a starkly reduced rendition of Classical Arabic grammar, resulting from an unusually extreme interruption in transmission. We do not intend to conflate the new Arabics with this degree of reduction: rather, Nubi represents a pole on one end of a cline, upon which the new Arabics occupy an intermediate point.

Nubi emerged as a pidgin Arabic in southern Sudan in trade settlements, first established in 1854 and controlled by Egyptians commanding massive numbers of soldier-slaves recruited from among local Africans speaking Nilo-Saharan and Niger-Congo languages such as Dinka, Bari, Mamvu, Zande, and Lumbara. Southern traders rebelled against the Egyptians amid the Mahdist movement, and their victory in 1882 left these southern settlements cut off from Egypt. The British soon relocated these soldier-slaves to assist in controlling Uganda, and eventually the pidgin Arabic creolized into a full language. This occurred, however, without Arabic as a target. English and Swahili were the new superstrate languages, and thus Nubi did not take its place in a diglossic relationship with the standard as so many other daughters of Classical Arabic did. Nubi became a natural language mostly via recruitment of its own internal resources, such as via grammaticalizations and reanalyses of Arabic material and appropriations of substrate features.

It would appear certain that Nubi originated in a pidgin rather than from waves of "approximation" of Arabic or "feature selection" between Arabic, Nilo-Saharan, and Niger-Congo. A modern reflex of the original pidgin persisted in southern Sudan as Juba Arabic (although this too has been adopted as a native language since the latter decades of the twentieth century).



Nubi's simplification of Arabic—be this Classical Arabic or the actual colloquial variety Nubi's creators were exposed to (this either Egyptian or, as Owens argues, Sudanese)—is massive (Owens 1997). The phonemic inventory is distinctly unmarked, devoid of emphatics or pharyngeals and with only one length distinction in vowels. Nonconcatenative morphology has disappeared; compounding is central in nominal derivation, while tense and aspect are marked with preverbal prefixes. The construct possessive is absent, possession connoted by an invariant particle. Suppletive plural forms, remnants of broken plurals, are few in number, plural usually expressed with a suffix *-á*. In general, in its low inflection and degree of elaborification, Nubi grammar bears the hallmarks of a language born amid rapid non-native acquisition by adults in a context where the lexifier language was only distantly available, including ample substratal transfer such as tone (although only minimally contrastive lexically or morphosyntactically) and vowel harmony.

This, then, is what happened to Arabic under conditions uniquely ill suited to anything approaching full mastery. From this, it follows that in situations where acquisition was blocked to a lesser but significant degree, we would expect lesser but significant abbreviations of Arabic structure. In this light, Arabic, as harbored by groups traditionally impervious to outsiders—that is, the Bedouins—will hew closest to Classical, while Arabic as acquired by massive numbers of adult foreigners will stray from this to a degree intermediate between the Bedouin varieties and the Nubi Creole extreme.

Thus, Classical Arabic represents “normal” language transmission, a product of language mutation amid small, isolated groups, with the high degree of elaboration, almost counterintuitive to an English speaker, typical of such languages. The sedentary Arabics resulted when Arabic was successfully acquired but structurally reduced by foreign learners across the new Islamic Empire, and hence their status as “Mandarins” compared to Classical Arabic. Bedouin Arabics have developed and survived under conditions inhospitable to non-native acquisition, and hence their intermediate position between Classical and sedentary Arabic. Pointedly, Bedouin varieties are most conservative in the Arabian heartland (Ingham 1992).

### 6.3. *Arabic Inflection along the Cline*

Kusters's (2003) analysis of the fate of inflectional morphology in Najdi, Moroccan, and Nubi Creole Arabic neatly confirms this account.

Under Kusters's analysis, Najdi Arabic remains approximately at par with Classical Arabic in terms of complexity of verb inflection (*ibid.*, 131–35). In terms of his Economy principle (roughly paralleling what I term as overspecification), Najdi has lost dual marking and mood marking, albeit compensating to a considerable degree for the latter with lightly conventionalized preverbal markers.<sup>5</sup> Najdi increases Classical Arabic's homonymy quotient in conflating active and passive verb forms in cases where the intransitive active imperfect's first vowel changes from *a* to *i*, thus mirroring the passive form. Moreover, new allomorphy is introduced in that this change does not occur with transitive verbs, and in some subdialects applies only to the first two persons. While some Classical Arabic allomorphies are eliminated by the weakened distinction between *i* and *u*, Najdi's object suffixes are subject to phonologi-

cally conditioned allomorphy to an extent beyond their Classical Arabic sources, in this recalling, in type though not extent, the Aramaic and South Arabian situations: *kital* “he killed,” *ktiluh* “he killed him,” *ktalan* “he killed her” (ibid., 129).

Moroccan Arabic, representing the sedentary varieties, is quite different under Kusters’s analysis (ibid., 142–46). There is even more Economy (i.e., less overspecification): in addition to the Najdi losses there is that of gender marking in the second- and third-person plural. Where Measures Five through Eight connoted various and only partially predictable shadings of passive, reflexive, and reciprocal meanings, Moroccan retains only two of these measures and connotes the valence with a single suffix *-t* of relatively unselective lexical application. Moroccan has less fission than its ancestor: stem-internal vowel distinctions distinguish aspect regularly only in weak verbs, leaving the affixes to carry most of the load (and not allomorphized by the internal vowels as in Classical). Another case where Moroccan decreases allomorphy, in its treatment of weak verbs (specifically those with *w* as a consonant) was described at the end of 3.4. In broad terms, Kusters’s analysis confirms that in its verbal morphology, Moroccan is simpler than Classical.

Then Nubi Creole Arabic has few inflectional affixes at all, and nonconcatenative verb morphology does not exist. What prefixes do exist are invariant for person and number, and only lightly allomorphized via phonology. The causative is expressed by an invariant suffix (borrowed from Swahili), and the passive is expressed with stress and high tone (Nubi’s creators spoke tonal languages) on verbs’ final vowel, applicable to intransitives as well as transitives (ibid., 149). Clearly this grammar will not submit to an analysis focusing on the ways in which simplifications can yield complexities. This is, *tout court*, a striking breakdown of an initially complex system of inflection, with much of its remnant complexity having been filled in by strategies imported from the substrate languages (Nubi also has some nominal derivation expressed with tone).

## 7. Two Excursuses

### 7.1. *The Bedouin Question*

A question has been left unanswered: why are the Bedouin dialects not even more conservative and innovative than they are? Under normal conditions, we would expect them to be as elaborated as Aramaic and South Arabian, with fewer losses and more innovations. Yet as Versteegh (1997, 141) notes, the Bedouin dialects for all of their conservativity are “clearly of the New Arabic type.”

This was especially highlighted in much of the response to Ferguson’s (1959) claim that his koine features distinguished the sedentary varieties from the Bedouin ones, according to a scenario in which the koine developed in military camps while the nomadic Bedouins retained the original language. Blanc (1970) conclusively showed that Bedouin varieties shared the dual loss, and future work has shown that they share many more of the koine features and others with the sedentary varieties, especially losses.

A natural assumption is that interactions between the Bedouins and settled populations drew the Bedouin varieties somewhat away from their conservative template.

Versteegh (1997, 141) assumes that “not even the Bedouin could escape the effects of sedentary civilization.” But this explanation is less clear when viewed at closer hand.

If the fine-tuning of language contact theory has taught us anything over the past twenty years, it is that mere contact between peoples does not guarantee grammatical transfer. There has been essentially no in-migration into Bedouin communities by sedentary Arabs. The Bedouins’ relationship to sedentary Arabs has always been that of the trader passing through or spending a summer to graze animals. In addition, Bedouins view themselves as the “purest” strain of Arab, and the sedentary varieties differ from Bedouin ones most saliently in absences—that is, to the untutored hearer, gaps, “mistakes.”

Just what, then, would cause Bedouins to incorporate into their casual speech “mistakes” made by out-group members encountered on a piecemeal basis, and incorporate them so thoroughly that they become decisive input to their children? Lexical transfer is much more natural under such circumstances, and as such, sedentary Arabics have lent a great deal of words to Bedouin varieties. But the available historical record supports neither heavy acquisition of Bedouin Arabics by outsiders nor Bedouins valuing sedentary speech and imitating it. Some might suppose that the Bedouins’ pride in their dialects’ closeness to the Classical language of their pre-Islamic poetry might discourage language change (e.g., Kusters 2003, 101). This type of point is essentially unfalsifiable, but in any case, it still leaves the question as to why there has been so much loss nonetheless.

We seek longer-term and more intimate interactions than trade or summer breaks. Language contact accounts that attribute major transfer to interactions of this kind rarely address human reality. The standard English-speaking reader is asked to imagine how likely he or she would be to come away from a summer in Scotland or Jamaica having internalized not only a few words, but aspects of the phonology, morphology, and syntax of the regional dialect. For those “language sponges” who experience linguistic “hangovers” of a week or so after travel of this kind, especially imagine retaining any of these features for longer than that week after returning home—and to such a degree that your children grow up speaking like the people you vacationed with! Obviously this is unimaginable, even if you spent every summer in the area. Perhaps one’s young children, with their absorptive abilities, would come away from the summer vacations having picked up even more shards of the speech style—but spending nine months of each year with their parents and local peers would erase this quickly.

This particular question about Bedouin speech appears not to have been addressed previously, and historiography offers only glimmers as to why Bedouins have not spoken something even closer to Classical Arabic (or more properly one of the Arabic dialects that existed before the formal codification of Classical Arabic). For some Bedouin groups, most contact with sedentary peoples is with ex-Bedouins or settled members of the same culture, such as in the Najd (Ingham 1984). But many others have had extensive contact with speakers of sedentary Arabics, as in Syria and Lebanon. Indicatively, today Bedouin dialects of the latter type are less conservative (Rosenhouse 1984, 260). As Rosenhouse has it (263), “Bedouin dialects adapt to their neighborhood.”

However, accounts of this kind generally describe exchanges of lexical items and free grammatical morphemes, and refer to modern developments such as Bedouin dialects being diluted by the constant exposure to other Arabics after the communications revolution, or Bedouins gradually giving up their lifestyles. They are of little use in revealing how much contact earlier speakers of a type of Classical Arabic, while still fiercely committed to a nomadic lifestyle not yet threatened by globalization, would have had with sedentary speakers sustained and intimate enough to lead them to shed their case markers or feminine comparative forms.

An honest approach to the problem within the bounds of my thesis is to hope that further research on Bedouin history can shed some light on the issue, if the information is not lost altogether. Perhaps there were significant sedentary incursions into early Bedouin groups, before migrations out of Arabia (which would explain why all of the modern varieties are of the new Arabic type). In this, the large numbers of Africans brought to Arabia as slaves just may have been a source of reduction. However, to confirm this, we would need to uncover more evidence than the anecdotal characterization of a single Bedouin group as racially mixed as that that Kusters (2003, 101ff.) usefully refers to. Or, perhaps Bedouins have had longer-term and more intimate contact with sedentary groups than generally supposed.

For now, however, the fact remains that the new Arabic dialects with the least exposure to non-native acquisition are the most elaborated. This squares neatly with the goals of the monograph. The question is, specifically, why Bedouin dialects are not *more* elaborated—a puzzle, but not a contradiction.

## 7.2. *Pidgin Arabic Precursors?*

Versteegh (1984) supposes that the new Arabics began as pidgin Arabics, and were later brought closer to Classical Arabic because of prescriptive influence. But if by this Versteegh means pidgins with a degree of conventionalized rules and of extended lifespan (analogous to pidgins such as Fanakalo and Chinook Jargon), I would submit that the best position here would be agnostic.

Traits in a grammar similar to those commonly found in pidgins and creoles do not necessarily entail that the grammar was born as a pidgin or creole, and are equally analyzable, as Kusters (2003, 156) aptly notes, as results of imperfect second-language acquisition. It would be surprising if in the regions Arabic spread to, there were not typically stages when most non-native speakers only controlled a pidgin-level competence—acquisition of a new language has to begin somewhere, especially under untutored conditions. The pidgin Arabic briefly documented in eleventh-century Mauritania by Thomason and Elgibali (1986) would be a “snapshot” of such a stage. But presumably, in contexts where Arabic was widely available as a model, many learners surpassed the pidgin stage quickly depending on their life experiences, while new generations were in a position to acquire much fuller competence, even if the large number of non-native speakers lent them a version of Arabic abbreviated to enough of a degree to create this chapter’s figure 7.1.

In this light, there are several contexts in the language contact literature in which this pidgin stage would appear to have been a brief, transitory one in which a merely

abbreviated, rather than pidginized, rendition of the target language emerged quite rapidly. The search for a pidgin or creole antecedent to Popular Brazilian Portuguese has been futile. The semicreole Réunionnais French also seems to have quickly established itself more or less as it is today, despite Chaudenson's (1979, 1992) account of its beginning as an importation of Mauritian Creole (Baker and Corne 1982; McWhorter 2000b, 187–90). African American Vernacular English may well have emerged comprising a now extinct semicreole pole, but documentary evidence remains opposed to earlier claims that it grew from a pidgin or Gullah creole (McWhorter 2001).

There is evidence that the new Arabics were similar cases. As Diem (1978) and Kusters (2003, 156) note, if Islamic identity pulled speakers' pidgin rendition closer to the standard, then Judaic Arabic varieties would be further from it, when in fact they are not. Moreover, the longevity of Juba Arabic pidgin in Sudan need not be treated as a modern parallel to other Arabophone contexts. This pidgin has remained in its reduced state for so long because Arabic has had merely adstratal, utilitarian status in its region for so long, while local languages have persisted as everyday varieties.

## 8. Conclusion

The difference between Classical Arabic and the new Arabics cannot be reduced to mere "drift," an analysis that founders upon the absence of, for example, any Algonquian or Australian language that has "drifted" in so abbreviatory a fashion over time. Meanwhile, the fact that the new Arabics are traceable to an initial koine, and that the varieties bear evidence of contact due to later population movements, leaves unexplained why what all of the varieties share is so strikingly reductive in comparison to Classical Arabic. The new Arabics bear hallmarks of the fate of natural language grammar amid the rapid and massive population movements that have only been possible in post-Neolithic civilizations.

# Malay

## *“A More Approachable Structure”*

### 1. Introduction

“A lingua franca needs to be easy to grasp, and Malay has a more approachable structure than its relatives,” notes Dalby (1998, 391). In light of the previous chapters, we might expect just this, as Malay<sup>1</sup> has been a principal lingua franca of the islands of Southeast Asia for a millennium at least.

But Dalby makes his remark in passing, and in a source aimed at laymen. In the Austronesianist literature, the issue of Malay’s relative structural simplicity attracts only the rarest of notice, and is never sustained. This contrasts sharply with the variant but robust degrees of attention paid to the equivalent issue among Semiticists, Sinologists, Iranicists, and specialists in the history of English.

Yet Malay is as anomalously decomplexified compared to its relatives as colloquial Arabic, Mandarin, Persian, and English are in comparison to theirs. As always, a natural first impression is that Malay’s grammar is the result of ordinary processes of language change, with its homology a perfectly ordinary fate out of many that the Proto-Austronesian rootstock has undergone over the past six millennia. But a more global perspective renders this assessment hard to support. Not only in comparison to its close relatives but also even to the world’s languages as a whole, Malay has a grammar more peculiar than is generally realized.

### 2. Malay and Austronesian

Austronesian divides into four subfamilies. Three comprise the roughly two-dozen languages of Formosa, while the remaining family, Malayo-Polynesian, comprises almost all of the one thousand or so languages in Austronesian. Obviously, a careful argument that Malay’s development has been anomalous in comparison to its

close relatives will require focus on a smaller subgrouping within Malayo-Polynesian.

Malayo-Polynesian is subdivided into three branches. The Eastern branch includes Oceanic and about forty-five languages of Irian Jaya and Halmahera Island just westward of it. The Central branch includes about fifty languages of the Lesser Sunda Islands (the chain starting east of Java with Bali and ending in Timor) and the Moluccas (of which Halmahera is one). Western Malayo-Polynesian comprises all of the others, covering the Philippines, Borneo, Sulawesi, Sumatra, Java, Malaysia, and two outliers of Micronesia, Chamorro and Palauan.

But even this subgroup includes hundreds of languages forming about thirty groups. This is too vast an array for the argumentational format most useful to this monograph, in which I need to address with close engagement the contrasting developments between closely related languages, in order to highlight that a particular language's evolution has departed from a general pattern too sharply to be accidental.

Determining what smaller division to address requires a judgment call, as the charting of Western Malayo-Polynesian taxonomy is not very advanced, beyond the identification of the thirty clusters, of currently uncertain relationship to one another (Clark 1990, 905).

Malay is classified as a member of the Malayic group. But Malayic is in large part a dialect complex of varieties of Malay (such as Kerinci and Banjar). There are only a few other Malayic languages, such as Minangkabau and Iban, and even these are so close to Malay that their status as separate languages has not been an unequivocal matter to all observers. An argument that Malay's evolution has been peculiar requires a larger, and optimally more diverse, data set than this.

The next level up would be a proposed group including Malayic and the mere three languages of Java: Javanese, Sundanese, and Madurese. This would be Dyen's (1965) "Javo-Sumatran hesion," traced to a "Proto-Malayo-Javanic" ancestor by Nothofer (1975). This would allow a slightly larger data set—but still one referring to fewer languages than in any other chapter of this book.<sup>2</sup>

In fact, the anomaly of Malay is most effectively revealed in its contrast with languages not only within but also beyond the compact region where Malaysia and Java intersect with Sumatra. However, higher branching relationships within Western Malayo-Polynesian have not yet been reconstructed (in fact, Western Malayo-Polynesian itself does not refer to languages that trace back to a single ancestor, but to a grab bag of languages that are neither Central nor Eastern Malayo-Polynesian, the latter two considered to trace to proto-languages [Blust 1999]).

It would seem that a rough but useful subdivision, traditional in the literature, between "Philippines-type" the tradition is "Philippines" and "Indonesian-type" languages will be the best guide to where to focus the discussion in this chapter. Philippines languages like Tagalog have their well-known elaborate focus systems and are mostly genetically close. The "Indonesian-type" (henceforth IT) classification refers not only to languages with grammars similar to Malay but also to ones more inflectionally elaborated (and obscure) of Sumatra, Borneo, Sulawesi, and other islands. Thus this chapter will examine Malay within the IT group.

### 3. Malay and the Indonesian Typology

To anyone familiar with Malay, a first response to the idea that the language is less complex than we would expect might be to remark upon Malay's healthy paradigm of derivational affixes and its two voice-marking inflections. Indeed, these are often the main course in Malay descriptions: as tones are to Chinese, verbal measures are to Arabic, and inflectional paradigms are to Latin, lengthy descriptions of the functions of its mostly derivational affixes are central to descriptions of Malay.

There is no doubting that there is nothing of the "streamlined" about these affixes. The prefix *per-* and the suffixes *-kan* and *-i* compete in creating transitive verbs: *tuhan* "God," *per-tuhan* "deify"; *basah* "wet," *basahkan* "to make wet"; *air* "water," *airi* "irrigate." In many cases the prefix occurs with either of the suffixes in circumfix style, as in *ladang* "unirrigated field," *perladangkan* "to open up (land) for cultivation." Elsewhere, the choice between prefix and suffix connotes fine semantic differences: *panjang* "long," *perpanjang* "extend the validity of," *panjangkan* "lengthen."

The semantic contribution of the morphemes tends to spread across various meanings not readily relatable semantically. The prefix *ter-* connotes the accidental (*duduk* "sit down," *terduduk* "fall on one's backside"); resultatives (*hormati* "respect," *terhormat* "respected"); and marks negative potentiality as in *makan* "eat," *tidak termakan* "not edible" (data from Sneddon 1986). Lexicalizations are common, and just as often the affixes occur in fossilized fashion, making no semantic contribution except to indicate verbhood.

Morphophonemics complicate matters further. The final nasal of one prefix that creates dynamic intransitives assimilates to voiced initial stops, replaces voiceless initial stops with a nasal in the same place of articulation, occurs as [nj] before [s], [ng] before vowels, and elides elsewhere (data from Sneddon 1996, 910):

TABLE 8.1. Morphophonemics of Malay *meN-*

meN + ajar	>	mengajar "teach"
meN + beli	>	membeli "buy"
meN + pakai	>	memakai "use"
meN + dengar	>	mendengar "hear"
meN + tulis	>	menulis "write"
meN + sewa	>	menyewa "rent"
meN + ganggu	>	mengganggu "bother"
meN + lihat	>	melihat "see"
meN + nanti	>	menanti "wait for"

Also, second cycles of derivation generally elide derivational morphemes from the first cycle: *obat* "medicine," *obati* "to treat medically," but *obatkan* "to have treated" and *pengobatan* "medical treatment."

Malay also has two verbal inflectional prefixes marking a sentence as subject-oriented (below, abbreviated as SO) or object-oriented (OO):



- (1) (a) Ali **memukul** Zainal kemarin.  
 Ali SO.strike Zainal yesterday  
 'Ali struck Zainal yesterday.'
- (b) Zainal **di-pukul** Ali kemarin.  
 Zainal OO-strike Ali yesterday  
 'Zainal was struck by Ali yesterday.' (Prentice 1987, 933)

Superficially resembling a typical active/passive distinction, this alternation is in fact linked to constraints such as definiteness, and the object-oriented marking is used much more frequently than a passive construction typically is, functioning in as much a pragmatic as grammatical sense. In Malay, the use of the object-oriented construction is conditioned by a far less emphatic highlighting of the patient than is necessary to spur a passive construction in English.

Thus it must be clear that there is no claim that Malay lacks complexity. The derivation, in particular, replete with degrees of desuetude, noncompositionality, fossilization, robustly distortional morphophonemic processes, and functional overlaps between morphemes, gives all indications of a system that has been in living use for a very long time, subject to the ravages of millennia.

This is, certainly, Malay's most apparent distinguishing feature in the cross-linguistic sense, a challenge to process for speakers of most other languages. But a question we might ask is: what else is there in Malay that creates an impression of overspecification, structural elaboration, or irregularity to a robust degree in the cross-linguistic sense?

The question is important because amid its relatives, Malay is by no means unique in its derivational apparatus. Rather, similar systems, including cognate morphemes, are a commonplace in IT languages. Malay is indeed unique—but in being the only one of its relatives in which so little else besides derivation and an inflectionally encoded voice distinction indicates its age.

#### 4. Malay versus Tukang Besi

A preliminary demonstration of the anomaly of Malay is a comparison with another representative of the Indonesian "type," Tukang Besi of Southeast Sulawesi. In terms of overspecification, structural elaboration, and irregularity, Malay is by no means typical of the IT group of which it is so often treated as representative.

In this section, page numbers for Tukang Besi data refer to Donohue (1999), and page numbers for Malay data refer to Sneddon (1996) unless otherwise indicated.

##### 4.1. *Phonemic Inventory*

Malay typifies a tendency in Austronesian toward smallish phonemic inventories. It has twenty-two consonants, none marked in the cross-linguistic sense. It has eight vowels (Prentice 1987, 917).

Tukang Besi has only five vowels, but twenty-seven consonants, including a bilabial fricative, prenasalized voiced and voiceless segments, and implosives (15–16).

#### 4.2. Infixation

In Malay, infixation is not productive, *-em-*, *-el-*, and *-er-* occurring only in a limited number of words with lexicalized meanings (*tunjuk* “point,” *telunjuk* “index finger”) or with meanings identical to the root without the infix (*suling*, *seruling* “flute”) (25).

Tukang Besi has a grammatically central infix *-um-* which marks future and irrealis when focus on the subject is intended (154–56) (for meaning of gloss “3I,” see 4.6):

- (2) Na-b-**um**-aiara  
 3I- pay.IRR  
 ‘She will/wants to pay.’ (155)

In this, *Tukang Besi*’s morphology surpasses Malay’s in complexity, in that infixation contravenes, for example, the Isomorphy Principle as outlined by Kusters (2003, 30–34), stipulating that grammars tend to order morphological items according to how cognitively relevant they are to the stem. For example, Bybee (1985) proposes that grammars hone toward an ideal morphological ordering as following: stem—valency—voice—aspect—tense—mood—number—person—gender.

As such, *Tukang Besi*’s marking of mood within the stem, rather than toward the outer edge of a sequence of prefixes or suffixes, qualifies as a flouting of ideal processibility. Of course, languages commonly exhibit wrinkles of this kind. But here, *Tukang Besi* does so in a fashion that Malay does not, surpassing it in what under my metric qualifies as a combination of structural elaboration and irregularity.

#### 4.3. Determiners

Malay has no definite article; *seorang* occurs optionally as an indefinite article, but does not approach the grammaticalized nature of, for example, English *a*. Malay can be said to mark definiteness in that subjects are required to be definite, such that to indicate an indefinite subject, the verb must be marked with the object-oriented affix and the object promoted to subject position:

- (3) Anak Ali di-gigit anjing.  
 child Ali OO-bite dog  
 ‘A dog has bitten Ali’s child.’ (Prentice 1987, 934)

But even here, the indefiniteness of the notional subject is left zero-marked. Unless special explicitness is desired, overall Malay is a language that tends strongly to leave the distinction between definiteness and indefiniteness to context.

One might attribute this to an areal tendency of the notoriously telegraphic grammars typical of the Southeast Asian Sprachbund. But then *Tukang Besi* obligatorily marks nouns for referentiality, and not with just a single item. One, *na*, marks

referentiality and another, *te*, occurs elsewhere, marking a distinction that articles tend to cover along with definiteness and indefiniteness in most Indo-European languages (63–65):

- (4) No-‘awa-‘e    **na**        boku **te**        ana.  
3R-get-3OBJ NOM book CORE child  
‘The child got the book.’ (161)

4.4. *Demonstratives*

Malay demonstratives denote a simple proximal/distal distinction with *ini* and *itu* (129).

Tukang Besi, like most IT languages, marks a three-way distinction in its demonstratives, between proximal, distal, and distant. In addition, *Tukang Besi* distributes this distinction across two additional paradigms, one indicating referentiality and another “presentative” one used when explicitly and physically presenting an object (137):

TABLE 8.2. Demonstratives  
in *Tukang Besi*

	actual	referential	presentative
this	ana	meana’e	kaana’e
that	atu	meatu’e	kaatu’e
yonder	iso	measo’e	kaaso’e

4.5. *Classifiers*

Modern Malay uses three numeral classifiers, *orang* for people, *ekor* for living creatures, and *buah* for inanimates. The numeral “one,” *satu*, becomes the prefix *se-* before classifiers: *seorang guru* “one/a teacher” (134–35).

Donohue lists twelve classifiers in *Tukang Besi* and states that there are likely others (109). In addition to functional equivalents to *orang*, *ekor*, and *buah*, these include classifiers for small, long, foldable, and pointed objects; soap, social groups, transport, cloth, and limb extremities (109–10). Furthermore, more numbers than “one” alter phonetically when used before classifiers, sometimes significantly: *sa’asa* “one” > *sa-*, *gana* “four” > *hato*, *no’o* “six” > *nomo* (107). *Tukang Besi* surpasses Malay in overspecification and irregularity here.

4.6. *Tense-Aspect Marking*

In Malay, tense and aspect are not marked obligatorily. The following sentence can have several meanings.

- (5)    Dia pergi ke kantor.  
      he    go     to office (197)

It can mean “He goes to the office,” “He went to the office,” “He is going to the office,” “He will go to the office,” and so on, with the meaning recovered by context. Tense and aspect can be marked with free morphemes *sudah* (*Dia sudah tidur* “He has slept” [198]) for perfective, *sedang* for progressive, and *akan* for future.

In *Tukang Besi*, affixal marking is obligatory for distinctions such as perfective and progressive, the latter expressed according to a fine semantic distinction:

- (6) No-waliako-**mo** di kampo-no.  
 3R-return-PF OBL village-3POSS  
 ‘They have gone back to their village.’ (172)
- (7) No-homoru-**ho** te wurai.  
 3R-weave-yet CORE sarong  
 ‘She’s still weaving a sarong.’ (173)
- (8) No-homoru-**do** te wurai.  
 3R-weave-EMPH CORE sarong  
 ‘She’s now weaving a sarong (and will do this first before beginning the curtain cloth).’ (175)

Moreover, *Tukang Besi* makes an obligatory distinction in subject-indexed prefixes between realis and irrealis (except in the first-person singular). Thus the paradigm (113):

TABLE 8.3. Realis/Irrealis Pronominal Affixes in *Tukang Besi*

	tonic	realis	irrealis
1S	iaku	ku-	ku-
2S	iko’o	‘u-/nu-	ko-
3S	ia	no-/o-	na-/a-
1(paucal)	ikami	ko-	ka-
1PL	ikita	to-	ta-
2PL	ikomiu	i-	ki-
3PL	amai	no-/o-	na-/a-

Sentences therefore contrast as such, in which R = realis and I = irrealis:

- (9) No-wila legolego.  
 3R-go arms.swinging  
 ‘He was walking, swinging his arms.’ (130)
- (10) Na-bumaiara  
 3I-pay.IRR  
 ‘She will/wants to pay.’ (155)

*Tukang Besi* is again, then, more overspecified for mood than Malay. Malay, for example, has no equivalent, with bound or free morphemes, for the regular indexing of realis and irrealis in *Tukang Besi*.

4.7. *Pronominal Allomorphy according to Grammatical Role*

Malay pronouns occur in a paradigm of bound forms sensitive to person but not number (165):

TABLE 8.4. Pronominal Clitics in Malay

	free	clitic
1st	aku	ku-
2nd	(eng)kau, kamu	kau-, -mu
3rd	dia	-nya

These are used agentively as proclitics in object-oriented sentences, as object-marking enclitics, and as possessive-marking enclitics:

- (11) (a) Buku ini sudah **ku**-baca.  
book this PERF 1-read  
'I have read this book.' (166)
- (b) Narti menunggu-**ku**.  
Narti wait-3  
'Narti is waiting for me.' (165)
- (c) rumah-**ku**  
house-1  
'my house' (166)

There is complication, in that when marking objects, the clitics only occur with verbs prefixed with *meN*-:

- (12) (a) Dia mengambil kue itu lalu makan.  
he take cake that and eat  
'He took the cake and ate (it).'
- (b) Dia mengambil kue itu lalu **me**-makan-**nya**.  
he take cake that and SO-eat-3  
'He took the cake and ate it.' (165)

Also, when the third-person clitic is used to denote the subject in object-oriented sentences, the *di*- object-oriented inflection is maintained whereas it is omitted in the first and second persons (cf. 11a):

- (13) Narti **di**-tunggu-**nya**.  
Narti OO-wait-3  
'He is waiting for Narti.' (166)

However, the usage of these clitics is optional in all of the contexts: free pronouns are also grammatical, that is:

- (14) Buku ini sudah **aku** baca.  
 book this PERF 1 read  
 'I have read this book.' (166)

In *Tukang Besi*, first, pronominal affixes vary according to both person and number, in the latter case extending to the paucal in the first person. In addition, besides the realis/irrealis distinction for subject, there are separate paradigms not only for possessive and object but also another paradigm for dative pronouns (113).<sup>3</sup>

TABLE 8.5. Pronominal Affixes in *Tukang Besi*

	tonic	realis	irrealis	possessive	object	dative
1S	iaku	ku-	ku-	-su	-aku	-naku
2S	iko'o	'u-/nu-	ko-	-'u	-ko	-nso
3S	ia	no-/o-	na-/a-	-no	-e	-ne
1(paucal)	ikami	ko-	ka-	-mami	-kami	-nsami
1PL	ikita	to-	ta-	-nto	-kita	-nggita
2PL	ikomiu	i-	ki-	-miu	-komiu	-ngkomiu
3P	amai	no-/o-	na-/a-	-no	-e	(amai)

With its affixes indexed to seven person/number distinctions, of obligatory usage, and marking five rather than three grammatical categories, *Tukang Besi* vastly surpasses Malay in overspecification in the pronominal realm. Malay exhibits structural elaboration in licensing the appearance of the object-oriented inflection *di-* to the third person, and in the grammaticality of the affixes according to the presence of *meN-*, but then *Tukang Besi* further complexifies its usage of these affixes in the fashion indicated subsequently.

#### 4.8. Redundant Concord

In *Tukang Besi*, subjects are marked with concordial affixes on the verb obligatorily, and objects optionally but usually, leading to ordinary sentences reminiscent of polysynthetic languages:

- (15) **No-**'ita-**'e** na kene-no te ana.  
 3R-see-3OBJ NOM friend-3POSS CORE child  
 'The child saw its friend.' (51)

Malay displays nothing of the sort: pronominal affixes do not occur as markers of concord with full NP arguments.

Finally, it must be stressed that Malay's derivational machinery does not "balance it out" in comparison to the above elaborations in *Tukang Besi*. For example, *Tukang Besi*, too, has a healthy mechanism of derivational affixes marking the causative, passive, applicative, reciprocal, and other distinctions:

- (16) No-**to-** **pa-** ala- mo na iai-su te kau.  
 3R-PASS-CAUS-fetch-PERF NOM younger.sibling-my CORE wood  
 'My younger sibling was made to fetch some wood.' (297)

Note also that the equivalent Malay sentence would have no indication that the sentence fell under realis mood, and the indication of the perfective would be optional. Nor would there be overt marking of the nominativity of the notional subject.

#### 4.9. *Inalienable Possession*

Malay does not mark alienability; for example, the choice between free and bound pronominal in the possessive is not sensitive to this distinction.

But *Tukang Besi* does mark alienability. The marker *mai*, used alongside the bound possessive form, emphasizes inalienability:

- (17) O-mosega ala'a na ana-su **mai** iso la.  
 3R-naughty just NOM child-1S.POSS INAL yonder EMPH  
 'My kid is nothing but naughty!' (347)

Also, at the clausal level, alienability is distinguished with the usage of the verb *hoto* as opposed to a construction with an existential verb:

Inalienable

- (18) Ku-**hoto** wunua to'oge.  
 1S-have house big  
 'I have a big house.' (348)

Alienable

- (19) **Ane** ke ana-su.  
 exist and child-1S.POSS  
 'I have children.' (350)

### 5. The Incredible Lightness of Malay

*Tukang Besi* compared to Malay is rather like Icelandic compared to English, in which a language that retains much of a family's original equipment while having transformed other aspects of it into features equally complex contrasts with one in which loss has vastly predominated in its life cycle. However, there are similar contrasts of lesser but robust degree between Malay and all of its relatives for which substantial descriptions exist.

#### 5.1. *Available Data and Its Implications*

To be sure, the languages with substantial descriptions constitute a small subset of the IT languages, which number 377 according to the Summer Institute of Linguistics Ethnologue (the figure arrived at by subtracting the 154 Philippines languages it lists from its 531 total for Western Malayo-Polynesian). To date, vast numbers of the minority languages, many spoken by small, isolated groups, have been documented only in wordlists or very brief grammatical sketches. It is sobering to find, for example, that at this writing there does not yet exist a substantial grammatical description of Balinese.

However, books and articles by Austronesianists making brief reference to the lesser-documented languages give a consistent impression that there are no languages as streamlined as Malay in the IT group. That is, neither my research nor my communications with Austronesianists have shed light on an IT language (or languages) beyond Malayic that stands out as Malay does in this regard.

Instead, specialists regularly document a southward transformation from the focus-oriented typology of the Philippines-type languages to an equally overspecification and structurally elaborated voice-focused one in IT languages (cf. section 7). In this work, Malay and its close relatives are parsed as a final, and structurally abbreviated, step in the transformation. Moreover, as we will see, even Malay's close relatives retain a degree of overspecification, structural elaboration, and irregularity significantly absent in Malay itself.

### 5.2. *Malay and Its Relatives*

Hence my presentation of the contrast between Malay and other IT languages in figure 8.1.<sup>4</sup> Gray indicates that the language exhibits the feature with greater overspecification or structural elaboration than in Malay, rather than that Malay lacks the feature entirely. Thus white means not that Malay has no phonemes (!); rather, gray means that the language has cross-linguistically marked phonemes that Malay does not (which also always entails, in this data set, that it has a higher phoneme count in general). Malay has three numeral classifiers in common use, but the languages marked gray for classifiers make use of significantly more, and so on.

One IT language that has a solid description (Durie 1985) is not included in the figure, Acehnese of northern Sumatra. This is because it is an outlying member of the Chamic group of Western Malayo-Polynesian, which has a unique history. As presented in Thurgood (1999), Chamic languages are descendants of an Austronesian language that underwent especially heavy contact with Mon-Khmer languages in present-day Vietnam. In Vietnam, the result was primarily monosyllabic languages with either tone or analogous register distinctions (i.e., phonemic distinctions between “breathy,” “clear,” and “creaky” voice), paralleling Mon-Khmer so closely in typology that their membership in Austronesian was long considered questionable. Just why some Chamic speakers migrated to Sumatra is unclear. But Acehnese retains a degree of Mon-Khmer influence, lesser than its Vietnamese sisters, but robust enough that its structure is too divergent from Malay and other IT languages to serve usefully in the argument of this chapter, even though in terms of comparative lexical reconstruction Chamic happens to classify as a close relative of Malayic, forming with it a Malayo-Chamic level within Western Malayo-Polynesian.

Our interest is in what happened to Western Malayo-Polynesian languages during relatively direct transmission over the ages, as compared to what I propose happened to Malay. No language is immune from the effects of contact—for example, a great many Western Malayo-Polynesian languages have been deeply affected by Malay, especially on the lexical level, and many scholars have argued that aspects of modern Malaysian and Indonesian are the result of standardization based on Euro-



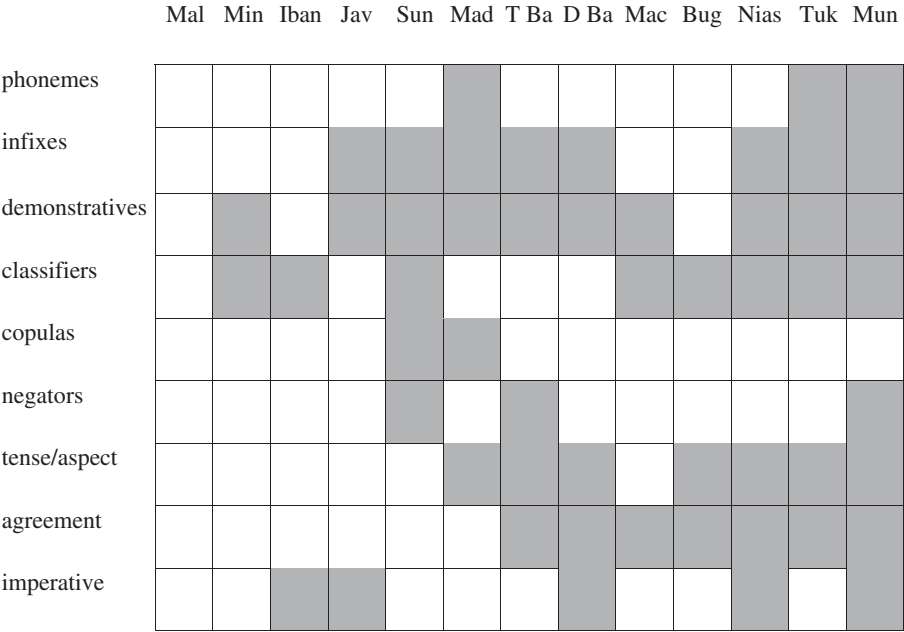


FIGURE 8.1. Overspecification and Complexity in Malay Compared to Its “Indonesian-type” Relatives

pean models. But Acehnese has undergone contact of an extremely transformative degree that renders its very group membership an academic challenge, such that comparing its development with Malay’s would be a matter of comparing apples and oranges (although, it is germane to note that it, too, would significantly surpass Malay in complexity as measured in the figure above).

5.3. *The Languages*

Minangkabau of Sumatra and Iban (aka Sea Dayak) of Borneo are fellow members with Malay of the Malayic subgroup, close enough to Malay to sometimes be considered dialects of it, the relationships in question roughly analogous to that between Spanish and Portuguese.

Javanese, Sundanese, and Madurese are the three languages of Java; Sundanese is spoken westward on the island and Madurese eastward. These three are grammatically similar to Malay, such that Dyen’s (1965) “Javo-Sumatran hesion” comprises languages akin to a degree reminiscent of the Romance languages.

Toba Batak and Daro Batak of Sumatra are fitfully mutually intelligible members of a dialect complex, such that “Batak” is a single “language” no more than is Italian or Hindi. Nias, a distant relative of the Batak group, is spoken on islands off of the western coast of Sumatra.

Macassarese and Buginese are spoken on the southwestern leg of Sulawesi, while Tukang Besi and Muna are spoken on islands off of the southeastern leg. The first two are more closely related to Malay than the latter two.

The languages in the figure comprise all of the IT languages for which substantial grammars exist at the time of this writing (a fact which alone makes a poignantly stark case for the urgency of documenting minority languages). However, henceforth I will make ample reference to less comprehensive, but useful, sources on other IT languages.

#### 5.4. *Features in the Figure*

Importantly, morphological loss is but one aspect of the losses in question. There is a general phenomenon of structural simplification to be explained.

*Phonemes.* Malay has no cross-linguistically marked phonemes. While smallish phonemic inventories are common in Austronesian, there are ample exceptions in IT, especially in the more isolated languages. Nias, for example, has a bilabial trill. In Sulawesi, Da'a has six prenasalized stops (/<sup>m</sup>p, <sup>m</sup>b, <sup>nt</sup>, <sup>nd</sup>, <sup>nj</sup>, <sup>kg</sup>/) (Barr 1995, 530); Wolio (Anceaux and Grimes 1995) and Uma (Martens 1995) have eight. As Trudgill (1996) notes in relation to Faroese, the proliferation of marked sounds is typical of languages rarely acquired by adults. Javanese and Madurese have retroflex consonant phonemes, which have often been assumed to be due to Sanskrit influence, but which Wolff (1988, 131–32) analyzes as internal developments from Proto-Austronesian.

*Infixes.* IT languages inherit infixation from their Philippines-type language precursors. No IT language retains infixation to anything approaching the Philippines-type languages, but they tend strongly to retain at least a single productive infix. Malay, however, has no productive infixation.

*Demonstratives.* Malay has a proximal/distal distinction in demonstratives, but this contrasts with the usual situation in IT languages, which is a three-way distinction including one for “yonder.” Some languages have an even finer gradation, such as the Batak varieties’ four-way one. The sensitivity to referentiality in Tukang Besi is not unique, such as in Muna: *bhai-ku aini* friend-my this “this friend of mine” (but not the others here/there) versus *bhai-ku ini* ‘my friend (here)/this friend of mine (already mentioned)’ (Van den Berg 1989, 92).

*Classifiers.* Modern Malay uses only three numeral classifiers in anything approaching regularity. The Batak varieties use just one more (hence marked with white boxes for this feature in figure 8.1), but IT languages that have classifiers commonly use more than this. Moussay (1981, 134–36) lists twenty in Minangkabau, for example.

*Copulas.* In Malay, there is no regularly expressed equative copula (although *adalah* and *ialah* can be used when subject or predicate are heavy in written language [Prentice 1987, 932]), while *ada* is used with existential and optionally with locative predications:

- (20) (a) Ayah guru.  
father teacher  
'Father is a teacher.' (Sneddon 1996, 237)
- (b) Tuhan ada.  
God exist  
'God exists.'
- (c) Ayah (ada) di kantor.  
father is LOC office  
'Father is in the office.' (ibid., 264)

This is a typical situation in IT languages, but some have a somewhat finer division of labor between copular morphemes. In Sundanese, for example, the *ada* cognate is ungrammatical when the predicate is of plural reference, constituting a kind of plural existential marking:

- (21) Di Jawa Barat (\*aya) loba pisan gunung.  
in Java west many very mountain  
'In West Java, there are very many mountains.' (Müller-Gotama 2001, 31)

Madurese has a negative existential copula (Davies 1999, 26); Muna has two grades of existential copula (Van den Berg 1989, 158–61).

*Negators.* Malay has a negator for the equative, one used for other predicates, and a negative imperative marker:

- (22) (a) Dia **bukan** guru.  
she NEG teacher  
'She isn't a teacher.'
- (b) Mereka **tidak** menolong kami.  
they NEG help us  
'They didn't help us.' (Sneddon 1996, 195)
- (c) **Jangan** tidur!  
NEG sleep  
'Don't go to sleep!' (ibid., 325)

This basic array of negators is found in many IT languages, but some elaborate it. Often this is to a moderate degree. Sundanese has a negative existential (Müller-Gotama 2001, 51–52). Minangkabau has an equative negator with a contrastive implication:

- (23) (a) Inyo **indak** panjojo lauak.  
he NEG vendor fish  
'He is not a fish vendor.'
- (b) Inyo **ukan** panjojo lauak.  
he NEG vendor fish  
'He is not a fish vendor (but something else).'

Elsewhere, the elaboration is sharply different from Malay. Muna has an equative negator, a general predicate negator, a predicate negator used for the future, a predicate negator used with participles, and a negative imperative marker:

- (24) (a) Anoa **suano**-mo guru-mani.  
 he NEG-PERF teacher-1PL.EXCL  
 'He is no longer our teacher.' (Van den Berg 1989, 212)
- (b) A Ntaapo-apo **miina** na-s-um-ampu.  
 ART Ntaapoapo NEG 3SI-go.down  
 'Ntaapoapo did not come down.' (ibid., 207)
- (c) **Pa** a-k-um-ala we sikola naewine.  
 FUT.NEG 1SI-go LOC school tomorrow  
 'I will not go to school tomorrow.' (ibid., 210)
- (d) Foo aini **pata** ni-uta-ku.  
 mango this NEG PPART-pick-my  
 'This is not a mango that I have picked.' (ibid., 211)
- (e) **Ko** mo-limpu itu  
 NEG IMP-forget that  
 'Don't forget that.' (ibid., 229)

*Tense/aspect marking.* As noted in 4.6, Malay does not mark tense and aspect obligatorily, often leaving it to context. But many IT languages mark tense and aspect regularly. Examples include Sumbawa, of the Lesser Sunda island of that name (Wouk 2000, 301, 304), and Madurese, which has inflections for the durative, habitual, and irrealis (Müller-Gotama 2001, 18–19). In Sumbawa's close relative Sasak (of Lombok island between Bali and Sumbawa), object-marking clitics are developing a distributional selectivity suggesting a movement toward inflectionhood, usually occurring only on verbs when they occur without an optional nasal prefix (Wouk 2000, 286–89).

*Agreement.* As noted in 4.9, Malay does not mark arguments with concordial affixes. This, however, is a loss unusual in IT languages as a whole. Some have only subject concord, but many also have object concord as well. Buginese is one example.

- (25) Təm-mu-íta-i kedona aruŋŋé  
 NEG-you-see-3PL gesture prince  
 'You won't see the gestures of the prince.' (Sirk 1979, 19)

In Muna, the subject prefixes are developed to the point of conjugational paradigms. There are three verb classes that take different person/number prefixes, correlating (roughly) with a distinction between transitive (*ae-lobhi* "I hit"), dynamic intransitive (*a-losa* "I emerge"), and stative intransitive (*ao-lowu* "I am drunk") (Van den Berg 1989, 54). In all three verb classes the prefixes vary for the realis/irrealis distinction also found in Tukang Besi (section 4.6), creating considerable allomorphy in some person/number classes (ibid., 57):

TABLE 8.6. Subject Prefix Allomorphy in Muna

	<i>ae</i> -class		<i>a</i> -class		<i>ao</i> -class	
	realis	irrealis	realis	irrealis	realis	irrealis
2S (polite)	te	tae	to	ta	to	tao
3S	ne	nae	no	na	no	nao

Meanwhile, various items exhibiting degrees of definiteness, such as definite objects, free pronouns, names, nouns with possessive suffixes, nouns with demonstratives, and even temporal adverbs, require that transitive verbs take the markers of the intransitive *a*-class:

- indefinite object:  
(26) **Ne**-rabu nuhua  
3SR-make pitcher  
'She is making a pitcher.' (ibid., 59)

- definite object:  
(27) **No**-rabu-e  
3SR-make-it  
'She is making it.' (ibid., 60)

*Imperative.* Malay does not mark the affirmative imperative with an affix or free morpheme. It uses either the bare verb, or subtractive marking in that with transitive verbs with definite objects, the *meN*- prefix is omitted:

- (28) Lihat (\*melihat) foto ini!  
look.at photo this  
'Look at this photo!' (Sneddon 1996, 324)

This trait is found elsewhere in IT such as in Toba Batak. But many other IT languages have more complex imperative marking. Even Malay's sister Iban transforms verbal prefix /bə/ to /pə/ in the imperative (a passive-marking alternation that the imperative is linked to): /bəjalayka/ 'to move (something)'; /pəjalayka/ 'move (that!)' (Omar 1981, 58). Javanese has imperative suffixes, which both cause alterations in the root and vary according to voice—the simple active suffix is *-ô*: *nulés*, *nulisô* 'write'; in the simple passive, it is *-en*: *tulés*, *tulisen* 'write (it)' (Horne 1961, 343). Imperative marking in Muna differs according to the verb classes—*ae*-class: **me-buri** 'write!'; *a*-class: **ø-kala** 'go!'; *ao*-class: **mo-lodo** 'sleep!' (Van den Berg 1989, 74).

6. The Tip of the Iceberg: Other Pathways  
of Austronesian Evolution versus Malay

As in the previous chapters, I have based the comparative table on broad contrasts amenable to counting. However, this captures only partially the unusually abbreviative nature of Malay compared to its relatives. Indonesian-type languages vary in the extent to which they surpass Malay in elaborification, structural elaboration, and irregularity along a continuum, starting with the the other Malayic varieties, and passing

through IT languages spoken near Malay in Sumatra, Java, and the Lesser Sunda Islands, and on to the languages of, for example, Sulawesi where Malay's streamlined structure is but a distant echo. Yet evolved traits absent in Malay are scattered throughout all of the other IT languages, all of them to my knowledge exhibiting at least some evolved traits—while crucially, Malay has no such features of its own.

### 6.1. Phonetic Evolution

In terms of phonetic evolution from Proto-Malayic, for example, Adelaar (1995, 435–37) notes that Minangkabau has ingrown sound changes that do not render it simply different from Malay, but further from the two languages' Proto-Malayic ancestor. This is what we would expect of a language that had evolved with less interruption than another.

In the Kerinci dialect of Malay, sound change has proceeded so far that all suffixes have been lost, but this loss left new complexity in its wake. Namely, the loss created distinct root forms for most nouns according to sentence-medial or sentence-final position: *binateng* “an animal,” *binatòn gědeng* “a big animal.” The rules for generating the sentence-medial form depend on vowel and preceding consonant, creating dozens of pairs of endings; for example, *ije/ijow* “green,” *panggèng/panggen* “call, invite,” *titay/titey* “cross a bridge” (Prentice and Usman 1978, 147–51). The distinction also becomes a new way to express features like definiteness and transitivity, as the morphemes indicating these distinctions preceded by a sentence-medial form become optional since the sound change now carries the functional load: *binateng minawng* “an animal is drinking,” *binatòn itoh minon nyo* [animal **that** drink **it**] “the animal is drinking it” > *binatòn minon* “the animal is drinking it” (ibid., 148). Then there are hundreds of words, however, that occur in only one form (ibid., 151), this class indexed to no predictable phonetic or semantic trait. Phonetic evolution in Kerinci has led to a more overspecified, structurally elaborated, and irregular grammar than standard Malay's.

Uma in Sulawesi is developing a register distinction between minimal pairs, in which a word-final glottal stop is giving way to pronunciation of the final vowel with creaky voice, this register marker even moving to final position when the word takes a derivational suffix, in the way that tones often behave. Uma (and other relatives with similar phenomena) may be on its way to developing pitch-based lexical contrasts (Quick 2002, 303–4). Malay exhibits nothing of the kind.

### 6.2. Morphophonemics

Morphophonemics in Malay are present (cf. section 3), but are not of an especially deep nature. They are of the sort that do not require much space to describe in grammars, and would be useful in elementary problem sets for beginning linguistics students. Other IT languages tend to have richer morphophonemics.

In Madurese, word-final voiceless consonants become aspirated when recast as onsets by suffixation (even though there is no grammarwide rule that onsets be aspirated), and the aspirated consonant also raises the following vowel according to a harmony pattern (Davies 1999, 10). The verbal suffix *-i* becomes *-an* before future suffix *-a*: *i-bhàrsi-i* “be cleaned,” *i-bhàrsi-an-a* “will be cleaned” (Stevens 1968, 83).

In Toba Batak, ten sandhi rules apply in ordered fashion across the grammar, conditioning more distance between the phonemic and the phonetic than in Malay: /álap hálak í/ > [álap pàlah í] “fetch them” (Woollams 1981, 59).

In Konjo of Sulawesi, before /h/, transitive prefix /aŋ/ surfaces as geminate *amm-*: /aŋ-huno/ > *ammuno* “to kill,” but then with vowel-initial roots, intransitive prefix /aʔ/ also becomes *amm-*: /aʔ-ore/ > *ammore* “cold (catarrh)” (Friberg 1995, 568–69). Thus *amm-* on the surface can emerge from two different underlying morphemes. In another Sulawesi language Wolio, the boundary between root and affix has been disrupted enough by sound changes to occasion outright irregularities, in which roots have forms that only occur before suffixes. Thus *kande* + *-aka* > *kandesaka*, *pili* + *-a* > *pilita*. In some cases, the special form only occurs before certain suffixes: *taqi* + *i* > *taqisi*, but *taqi* + *aka* > *taqiaka* (Anceaux 1952, 11).

Nias nouns, pronouns, and numbers occur in mutated forms when used as arguments of various kinds, such as patient, subject of intransitive verb, and others: *fakhe* > *vakhe* “rice,” *doi* > *ndroi* “thorn.” Basic mutations follow regular morphophonemic patterns, but there is a large class of vowel-initial nouns that take an irregular /n/, the mutated pronouns and numerals are irregular to the point of constituting a separate paradigm, and the mutation is rather particular in terms of where it happens not to occur (such as with various prepositions) (Brown 2001, 68–77, 341–71).

Malay is hardly Esperanto, but it does not present morphophonemic challenges of this kind. Deep morphophonemics are symptomatic of grammars that have been transforming without significant interruption for long periods, extreme examples being the famous mutations in Celtic, or the consonant gradations in Baltic-Finnic (e.g., Finnish and Estonian). That is, heavy use over time distorts sounds that occur in frequent conjunction, creating regular transformations that become productive grammar, with the distortion often continuing to the point of fossilizing into irregularity. Because allomorphy and irregularity challenge adult acquirers, non-native acquisition tends to undo this kind of “damage.” We would expect, then, that the IT language most widely acquired by adults over a millennium would have morphophonemics less rich than its relatives.

### 6.3. *Other Features*

On the subject of irregularity, while Malay usually leaves plurality unmarked and marks it where explicitness is desired via total reduplication, Javanese has some suppletive plural nouns, such as *wet* “tree,” *witwitan* “trees,” *gunoŋ* “mountain,” and *pagunungan* “mountains” (Horne 1961, 53). Javanese also has a “subjunctive” (hortatory, optative, prohibitive) suffix *-a* (Keeler 1984, 276–78).

As noted in 4.5, only the number “one” in Malay takes a different form before classifiers. *Tukang Besi* is hardly unique in IT in having more allomorphy of this kind. In Javanese, for example, *loro* “two” becomes *rong*, *telu* “three” becomes *telung*, *papat* “four” *patang*, *enem* “six” *nem*, and so on (Keeler 1984, 132).

Since ergativity is so common cross-linguistically, it would be unusual if it were absent in a family comprising a sixth of the world’s languages. Thus ergativity does occur in the family, and not only in the well-known examples in Polynesian, but in some IT languages. Uma of Sulawesi lacks *Tukang Besi* and Muna’s realis/irrealis

distinction in prefixes, but instead has ergative/absolute ones (Martens 1995, 543); Konjo (Friberg 1995, 570), Buginese (Abas and Grimes 1995), and Macassarese do as well. In Nias, mutation does not occur on agents, but does occur on other subjects and on patients. Not only can this be analyzed as rendering the language an ergative one but it also qualifies Nias as the only case known to date in which the absolute is marked and the ergative is not, rather than the typical reverse situation.

Malay uses *apa* ‘what’ as a general interrogative marker:

- (29) **Apa** dia sudah makan?  
 QU she PERF eat  
 ‘Has she eaten?’ (Sneddon 1996, 311)

But this is more common in writing and formal speech; its omission is grammatical, with intonation and context filling in. Daro Batak has a finer gradation of phrase-internal interrogative morphemes of this kind (Woollams 1996, 222). The *apa* equivalent is *kang*, with neutral interpretation; *kin* is used when an affirmative answer is expected, while *nge* and *ndia* in (31) express doubt:

- (30) Perjuma **kin** ia  
 farmer QU he  
 ‘Is he a farmer?’
- (31) Siat **nge** kari rumahna ah?  
 accommodate QU later house.his that  
 ‘Will his house be able to accommodate so many people?’

Certainly Malay can solicit affirmation and indicate doubt with tags, but for one, it encodes both distinctions with the same morpheme *bukan* (Sneddon 1996, 312), and in any case, Daro Batak also has tags. Overall, Daro Batak is more overspecified than Malay in marking shades of the interrogative.

#### 6.4. *The Issue of Degree*

Serendipity has it that Malay is not the least overspecified or structurally elaborated language of its group in categorical fashion. Malay has pronouns in all six person/number combinations plus an inclusive/exclusive distinction in the first-person plural typical of Austronesian. But Sundanese lacks a first-person plural pronoun; in Buginese, first-person pronouns are the only ones marked for plural. According to my sources, Javanese and Macassarese make no regular plural distinction in their pronouns. Javanese and Madurese do not have numeral classifiers (they are absent from grammars, and I have confirmed this gap with William Davies [March 2004 personal communication]). Iban has lost the *-i* transitivizing suffix that shares space with *-kan* (Omar 1981, 6–7).

But things like this are flutters in the data, within which Malay is overwhelmingly the least complex language according to the metric of this book. For example, despite its slight derivational discrepancy with Malay, Iban has dual-marked pronouns in all three numbers and all four persons (this including the inclusive/exclusive distinction in the first-person plural) (*ibid.*, 118).



This Iban trait is unusual in Austronesian as a whole, but, importantly, there appear to be no features in Malay that are superlative in overspecification, structural elaboration, or irregularity among its relatives. For example, the realis/irrealis distinction in many languages of Sulawesi is similarly exceptional in IT, as are scattered features like the high central yogh vowel in Sundanese (Müller-Gotama 2001, 6). But Malay has none of its own local divergent complexities analogous to *do*-support in English or the *bǎ* marker in Mandarin. Malay structure can be seen as representing a boiled-down prototype of IT, which all of the other IT languages not only reproduce but also elaborate to various extents.

## 7. From Protolanguage to Malay

Also in the diachronic aspect, in Malay we see a predominance of subtraction over addition and transformation—the hallmark of a grammar whose transmission has been interrupted to a considerable degree.

### 7.1. Verbal Morphology

Austronesianists chart the development of the “Indonesian-type” grammar as a gradual transformation of the “Philippines-type” grammars’ machinery, which obligatorily marks focus on a sentential constituent with both a trigger particle *ang* and affixal markers that vary according to the constituent in focus. Here is the process illustrated in Tagalog (Schachter 1990, 941) (AT = agent trigger, PT = patient trigger, DT = dative trigger, BT = benefactive trigger, TG = trigger marker, AC = actor, DR = directional):

(32) actor:

- (a) **Mag**-aalis                      **ang** tindero      ng bigas sa      sako para sa babae.  
 AT.PROG-take out TG storekeeper PT rice    DR sack BEN    woman  
 ‘The storekeeper will take some rice out of a/the sack for a/the woman.’

patient:

- (b) Aalisi-**n**                      ng tindero      **ang** bigas sa      sako para sa babae.  
 PROG.take out-PT AC storekeeper TG rice DR sack BEN woman  
 ‘A/the storekeeper will take the rice out of a/the sack for a/the woman.’

directional:

- (c) Aalis-**an**                      ng tindero      ng bigas **ang** sako para sa babae.  
 PROG.take out-DT AC storekeeper PT rice TG sack BEN woman  
 ‘A/the storekeeper will take some rice out of the sack for a/the woman.’

benefactive:

- (d) **Ipag**-aalis                      ng tindero      ng bigas sa      sako **ang** babae.  
 PROG.take out-DT AC storekeeper PT rice TG sack TG woman  
 ‘A/the storekeeper will take some rice out of the sack for a/the woman.’

Southward, this focus machinery gradually transforms into one marking voice (e.g., Blust 2002; Ross 2002), while bound pronominal forms emerge that, as they

often have cross-linguistically, lead to the development of conjugational paradigms (e.g., Van den Berg 1996; Wolff 1996). This, then, is the kind of homology seen above in languages like *Tukang Besi* and *Muna*.

But there is a crucial qualitative difference between the evolution from the Philippines type to the Indonesian type — roughly speaking, from Tagalog to *Tukang Besi* — and the evolution from the prototypical IT homology to Malay. Ross (2002, 49) presents the following reconstruction of verb morphology in Proto-Malayo-Polynesian, founded in a classic earlier reconstruction by Wolff (1973). Slashes between items indicate that they applied to the morpheme simultaneously, such as as a circumfix (e.g., R-/an) or as a prefix and infix (e.g., i-/in-); alternate markers are separated by commas:

TABLE 8.7. Verb Morphology in Proto-Malayo-Polynesian

	actor	patient	location	circumstantial
INDICATIVE				
neutral	-um-	-ən	-an	i-
perfective	-umin-	-in-	-in-/an	i-/in-
imperfective	-um-/R-	R-/ən	R-/an	i-R-
NON-INDICATIVE				
atemporal		-a	-i	-án
projective	-a	-aw	-ay	

Here is a busy system with overt marking linked to constituent class, this distinguishing the perfective, the imperfective, and the neutral, and distinguishing the indicative from the nonindicative.

Ross’s reconstruction of verb morphology in a later hypothetical IT language ancestor (2002, 53) is somewhat less overspecified: overt marking of the imperfective is lost. But then there is new sensitivity to voice (data rearranged from Ross’s original for optimal illustration of my presentation). In this table, *-ku*, the first-person singular pronominal clitic, stands for clitics in all persons and/or numbers. When *-ku* follows another morpheme, this is to indicate that the clitic occurred after said morpheme (e.g., *-an-ku*).

TABLE 8.8. Verb Morphology in a Hypothetical Early Indonesian-type Language

	patient undergoer	location undergoer	circumstantial undergoer
ACTIVE			
neutral	maN-, -um-	maN-/i-, -um-, -i	maN-/an
perfective	naN-	naN-/i	naN-/an
PASSIVE			
neutral	-ku	-i-ku	-an-ku
perfective	ku-, -in-/ku	ku-/i-, -in-/i-ku	ku-/an, -in-/an-ku

This is, overall, a reorganization of the Proto-Western Malayo Polynesian paradigm rather than an abbreviation of it. But it is harder to analyze the Malay descendant of this system in that fashion (data reorganized from Ross’s [2002, 52] presentation):

TABLE 8.9. Verb Morphology in Standard Indonesian

	patient object	location object	circumstantial object
ACTIVE	meN-	meN-/i	meN-/kan
PASSIVE	ku-	ku-/i	ku-/kan

The neutral/perfective distinction has collapsed, leaving less allomorphic overspecification in the active and less variation in morpheme placement in the passive; infixation is lost. There is one elaboration, in that in the third person in the passive, the suffix *-nya* appears along with prefix *di-* to express a pronominal actor (if no actor is expressed then *di-* occurs alone). But clearly this system is a considerable simplification of the one that Ross reconstructs for Proto-Malayo-Polynesian.

But IT languages, especially beyond the “Malayo-Javanic” realm, retain more distinctions than Malay, reflecting those marked in Ross’s reconstruction, as well as in the grammars of Philippines-type languages. In *Tukang Besi*, nominative marking is not linked to syntax (the nominative can be agent or object in transitive clauses) or semantics (any argument can be nominative); in this sentence, the instrumental is marked nominative:

- (33) No-tu’o-ke te kau kuu **na** **baliu-no**.  
3R-fell-3OBJ CORE wood ebony NOM axe-3POSS  
‘He felled the ebony tree with his axe.’ (Donohue 1999, 476)

Donohue argues that this is a reflex of the Philippines system’s obligatory marking of “trigger” constituents regardless of their grammatical role (*ibid.*, 53–54, 160–64, 488). Similarly, NP determination varies in *Tukang Besi* for purposes of disambiguation and monitoring status of participants (*ibid.*, 161; Mark Donohue, March 2004, personal communication). In (34a), *book* is more likely to be interpreted as indefinite and/or as new information when not marked as nominative:

- (34) (a) No-‘awa te boku **na** **ana**.  
3R-get CORE book NOM child  
‘The child got a book.’  
(b) No-‘awa-‘e **na** **boku** te ana.  
3R-get-3OBJ NOM book CORE child  
‘The child got the book.’

Compare in Tagalog:

- (35) (a) K-um-uha ng libro **ang** bata.  
get.AV GEN book NOM child  
‘The child got a book.’

- (b) K-in-uha ng bata    **ang**    libro.  
       get.OV GEN child NOM book  
       ‘The child got the book.’

Also, the typical IT language has bound morphemes obligatorily indicating tense and/or aspect, innovated as the perfective marking of table 8.8 wore away (cf. section 4.6). This contrasts with Malay’s loss of such marking, and only optional indication of the distinction with free morphemes.

Then we return to the development in IT languages of verb conjugations. This began with the development of bound subject markers, which is itself but one point along a cline toward fuller head-marking, such as the concordial marking of both subjects and objects in *Tukang Besi*. Even IT languages more closely related to Malay, such as *Macassarese*, have bound subject markers. Predictably, Malay is one of the IT languages that eschews this.

In the sequence of tables from Ross (2002), then, is another case that may seem at first glance “normal” because of the familiarity of the development of French from Latin. But the development of Romance involved a considerable degree of addition amid the subtraction. Romance lost most case-marking morphology on nouns—but then retained two grammatical genders and innovated determiners (these sensitive to gender and number); it lost passive-marking verbal paradigms, but developed conditional marking that Latin lacked. Grammars typically retain a certain degree of complexity. This renders the development from table 8.7 to table 8.9 rather striking, given that Malay has innovated no distinctions “compensating” for those lost in the tables. That is, standard Malay did not innovate ergative marking, or a fine-grained marking of shades of the past or future, or evidential marking, nor did it develop grammatical gender marking, or nominal roots varying for position as in *Kerinci Malay*.

And in this, we must extend our purview beyond Indonesia. Consider a bird’s-eye view of the world’s languages, in which the indigenous languages of the New World and Australia, Niger-Congo language families such as West Atlantic and Bantu, and the vast majority of Indo-European languages bristle with elaborate morphology marking a wide array of semantic niceties millennium after millennium, rather than shedding these features at the same time as Malay arrived at its current state.

### 7.2. *Proto-Malayic, Classical Malay, and Modern Malay*

But at this point, we have only seen how Malay developed from earlier stages in terms of overspecifications in the verb phrase. We desire a more comprehensive look at how Malay has developed.

Yet the grammar of Proto-Malayic, the proposed ancestor of Malay varieties and Malayic sisters *Minangkabau* and *Iban*, has to date only been reconstructed partially, in preliminary remarks on morphology and pronouns in Adelaar (1992), devoted mostly to lexical reconstruction. (Nothofer’s [1975] reconstruction of the next highest ancestor “Proto-Malayo-Javanic” is entirely devoted to lexical reconstructions.) Adelaar’s study, then, despite its keystone status, offers only glimmers of information for this chapter’s argument.

In terms of Malay's abbreviation over time, Adelaar reconstructs that Proto-Malayic had a three-way demonstrative distinction (Adelaar 1992, 127), anomalously lost in modern Malay. Then on the other hand, Adelaar also reconstructs that early IT infixes *-in-* and *-um-* were likely already fossilized in Proto-Malayic. However, he bases this conclusion on their fossilized state in the modern languages, when there might be no reason not to suppose that this fossilization has occurred in all of them *after* Proto-Malayic, in Malay because of heavy usage as a lingua franca, and in its sisters as the result of a wave effect due to massive bilingualism with Malay (there exists no Malayic language spoken beyond close contact with Malay itself).

Meanwhile, Classical Malay, despite its name, is properly not a stage of Malay from the mists of antiquity. While its documents preserve texts that mostly originated as far back as six hundred years ago, they represent copies made mostly as late as the eighteenth and nineteenth centuries. Especially because they were usually intended for oral performance, their language was modernized to a considerable degree. For this reason, there exist no Malay texts equivalent to *Beowulf* or the Qur'ān. The language of the Classical Malay manuscripts differs about as much from modern Malay as the language of Shakespeare does from modern English. (Classical Malay is close enough to modern Malay that dictionaries and grammars of "Malay" as late as the early twentieth century, before the standardization of modern "Indonesian," are based on Classical Malay, processed merely as a "high" variation on contemporary speech rather than as an archaically unfamiliar stage of the language.)

As such, Classical Malay is a language in which many of the losses exhibited in figure 8.1 have already happened. In the documents treated in Cumming (1991), for example, there are no concordial affixes, no morphological marking of the imperative (53–54), tense and aspect are indicated with free morphemes (69–71), and there are only proximal and distal demonstratives (23).

However, some losses have not occurred yet. Classical Malay uses a large battery of classifiers now marginal; Macdonald (1976, 83) lists a nonexhaustive fifteen moribund Malay classifiers, for example. Tadmor (2003) notes that the third-person pronoun was sensitive to a subject/oblique distinction in Classical Malay (*ia* versus *dia*). This was an overspecification now lost in comparison with not only IT languages with pronominal paradigms regularly sensitive to case, but ones closer to Malay like Madurese in which there nevertheless persists a genitive-marked *tang* "my" against *sengkoq* "I."

Meanwhile, Cumming (1991) addresses the fact that VS order is dominant in intransitive sentences in Classical Malay in contrast to modern Malay's default grammar-wide SVO order, such that Malay has lost a heterogeneous word order. Crucially, she ties this to the loss of grammatical-role marking items such as *akan*, which marked objects in Classical Malay but is now represented solely by the suffix *-kan*, now largely lexicalized with transitive verbs (i.e., the verbs usually cannot appear without it) and, in its productive uses, semantically evolved into related meanings such as causative, benefactive, and semantically unpredictable ones (*membacakan* can mean "read for someone," in the benefactive meaning, but also "to read aloud" [Sneddon 1996, 83]). Certainly the drifting of *akan* into boundedness and semantic ambiguity was a normal process of language change,

and it is reasonable to assume that SVO order is more likely in languages without morphemes regularly indicating grammatical role. But in broader view, at the Classical stage, Malay was already a language where only morphemes like *akan* performed this traffic-monitoring function — while in most IT languages, conjugational paradigms were in various stages of development, set to give overt signaling to subjecthood and often objecthood. Today, Malay retains inflections *meN-* and *di-* as pragmatic highlighters of actorhood and patienthood as a form of voice marking, but obligatory marking of grammatical role within the clause regardless of voice marking is unknown.

### 7.3. *Implications*

Obviously our knowledge of earlier stages of Malay is too sparse for this argument to proceed significantly on the basis of historical documentation to the extent that is possible for English, or even Chinese, Arabic, or Persian. Yet even the available sources on early Malay lead to a question: what overspecifications, structural elaborations, or irregularities has Malay innovated upon the IT template? The contrast between the development of modern Aramaic from Classical Aramaic, or even modern English from Old English, and what we know of earlier stages of Malay and its modern rendition, is quite striking. Compared to Adelaar's reconstruction of Proto-Malayic, for example, Iban has innovated a dual-marking pronominal paradigm over the plural-marking one Adelaar reconstructs (1992, 123). There are, to my knowledge, no significant grammatical distinctions that Malay can be said to have innovated upon its earlier stages.

To be sure, certain markers have developed within Malayic, such as the *di-* prefix and the evolution of *akan* from object marker to the more semantically lexicalized suffix *-kan* (Ross 2002, 55, based on Adelaar 1992). But *di-* merely reinforced an already established paradigm marking object orientation marked by clitics, joining in the third person the encliticized *-nya*; this created no new distinction. And *-kan* is more complex than its free morpheme source *akan* only in its variant degrees of fossilization upon roots, a mere eyedropper upon the blaze of such noncompositionality already rife in Malay derivation. Modern Malay has developed this mere peep of complexity since Classical Malay, a mere whisper amid a vast predominance of loss, such as of gradation of demonstratives and number of classifiers. Overall, under no analysis could we even begin to stipulate that modern Malay is either more or *even equally* complex overall, according the metric of this book, than it was at any earlier stage of its history.

In chapter 2 I noted that while a German speaker learning English is faced partly with variations upon familiar patterns (i.e., the specific range of prepositions), he or she also encounters a language where less must be attended to in terms not only of inflectional morphology but also a general degree of overspecification and structural elaboration. It would seem that IT speakers learning Malay have a similar experience, faced with 1) variations upon, and 2) abbreviations of, the familiar, but very little, or in many cases almost nothing, to be attended to that their own language leaves to context.

## 8. Malay in Actuality

The uniqueness of Malay, and the roots of this uniqueness in heavy usage as a second language, becomes especially clear in view of the fact that as used everyday by most of its speakers, Malay is even simpler than the standard variety—in this form, indisputably the least structurally complex language of the “Indonesian type.”

Standard Malaysian and Indonesian are, in actuality, artificial constructs drawn from earlier stages of Malay as represented in Classical Malay documents, promulgated by colonial rulers and indigenous leaders in the educational and formal contexts. The everyday language of most Malay speakers is sharply different from the Malay that they acquire in school or encounter in the media. Spoken Malay is a farago of dialects of limited mutual intelligibility, in a fashion paralleled, if not to the same degree, by the colloquial Arabics. These dialects are not standardized and are commonly considered “bad” Malay. Their social position is similar to that of Black English in the United States, complete with speakers’ reluctance to admit to using them, the tendency for speakers to switch to standard forms when asked for elicitation, and a paucity of formal descriptions.

8.1. *Nonstandard Malays*

Importantly, most of these colloquial varieties are not, like Kerinci or Banjar Malay, simply alternate renditions of Proto-Malayic as it happened to evolve in each region, along the lines of conservative Scandinavian dialects that retain all three Proto-Germanic genders, local dialects of Finnish, or even Minangkabau and Iban. Rather, like the new Arabics, most colloquial Malays are radically simplified in comparison to standard Malay.

Adelaar and Prentice (1996) characterize the dialects in question as descendants of the pidginized form of Malay used as a trade language, Bazaar Malay, terming them “Pidgin Malay Derived” (PMD) dialects. They list eight characteristics of these varieties, of which several are distinctly abbreviative of standard Malay:

1. Plural pronouns formed with singular pronouns + *\*orang* “human being”;
2. *\*ter-* and *\*ber-* as the only productive derivational morphemes retained;
3. causatives formed not with derivational affixes but with *\*kasi/\*beri* “to give” or *\*bikin/\*buat* “to make”;
4. *\*sama* or another word as a multifunctional preposition (also direct and indirect object).

Actually Adelaar and Prentice also list two features that qualify as substitutions for standard Malay items:

5. Reduced *\*pergi* as verb “to go” as well as preposition “to” (but standard Malay has preposition *ke* for “to”);
6. *ada* existential as progressive (but standard Malay has *sedang* for the progressive).

And there are two features that actually render the dialects more overspecified than standard Malay:

7. Possessive indicated with possessor + *punya* + possessed item (whereas stan-

dard Malay prefers simple juxtaposition, making only occasional use of possessive marker *dari* [Sneddon 1996, 145]);

8. reduced demonstratives *ini* and *itu* as determiners (whereas standard Malay has no determiners).

But descriptions of individual varieties of “PMD” Malays suggest further abbreviations. Grijns (1991) describes the Malay of Jakarta as having not only multifunctional *sama*, but no diphthongs, a neutralization of the standard Malay rule that allows only voiceless final stops, a virtual eclipse of subject-oriented prefix *meN-*, the collapse of transitivity suffixes *-kan* and *-i* into a single *-in*, the generalization of the *bukan* negator over *tidak*, and the absence of exclusive first-person plural pronoun *kami* (16–21).

## 8.2. Riau Indonesian

The same tendency is clear in the most comprehensively documented PMD dialect, Riau Indonesian of Sumatra, described and analyzed in several articles by David Gil. Gil describes Riau Indonesian as even more reduced from standard Malay than Prentice and Adelaar’s PMD prototype.

Riau Indonesian does not display any of the previous four nonsimplifying PMD features (depending on one’s grammatical characterization of the *punya* construction) (Gil 2001, 336–37). Instead, in a vast predominance of its aspects Riau Indonesian is a strikingly abbreviated rendition of Malay as presented in textbooks. It is essentially a pro-drop language. Its word order is free. There are no copulas, complementizers, or relativizers. There is no possessive marker according to Gil’s analysis. Tense and aspect marking are even less conventionalized than in the standard; and classifiers are rarely used (*ibid.*, 343–57).

Standard Malay’s agent-oriented *meN-* and object-oriented *di-* no longer encode anything resembling an active/passive voice distinction, but instead merely highlight actor and patient, respectively, and then are used only optionally, in a fashion straddling the line between syntax and pragmatics (Gil 2002). Thus whereas in standard Malay, *aku di-goreng* [I OO-fry] would mean “I was fried,” in Riau Indonesian it means “I fried it,” *di-* simply highlighting the implied patient. *Mister m-anggil*, in which *m-* is the morphophonemic rendition of Riau Indonesian’s *meN-* reflex *N-*, would in standard Malay mean “Mister calls somebody,” but in the context Gil transcribed it in meant “Call him ‘Mister’,” the *N-* highlighting the role of the actor (if I understand Gil correctly, translating roughly as “What you (should) call him is ‘Mister’”). In Riau, if passive meaning is explicitly intended, free morpheme *kena*, a verb meaning “undergo” is used:

- (36) Mister      kena      tembak.  
white.man    undergo shoot  
‘The white man got shot.’ (Gil, 2002)

Gil remarks that this radical deconstruction of standard Malay’s active/passive voice machinery is also found in various other nonstandard Malay varieties, and to a more extreme degree. Sulsel (southern Sulawesi) Indonesian eliminates *N-* en-



tirely. Colloquial Kuala Lumpur Malay lacks *di-* entirely and only retains *N-* marginally. The colloquial Indonesian of Irian Jaya lacks *N-* as well and also all but lacks *di-*, representing the end of a cline of eclipse of the use of these morphemes altogether.

This evolution is important in that this drift toward highlighting actorhood and patienthood, and then only optionally, qualifies not as just a variation upon, but a simplification of, standard Malay's use of these morphemes. In standard Malay, for one, the voice-marking system is obligatory rather than optional. In addition, it constrains word order, such that object-oriented verbs with *di-*, as equivalents of the passive in other languages, are followed by actors while subject-oriented verbs with *meN-* are followed by patients. Moreover, the voice marking in standard Malay conditions various rules regarding placement of bound pronominals (4.7), imperative marking (5.4), and *wh*-word movement, within which in object-oriented sentences, "what" can only occur relativized<sup>5</sup>:

- (37) *Siapa yang di-bicarakan di      rapat?*  
       what REL OO-discuss LOC meeting  
       'What was discussed at the meeting?' (Sneddon 1996, 316–17)

Finally, the passive is, in the Greenbergian sense of implicational universals, a marked choice of what to specify overtly and obligatorily. Countless languages leave passive interpretation largely to context (such as most creoles, none of which have morphemes dedicated exclusively to marking verbs or their arguments as passive except in levels influenced by Indo-European lexifier languages). It is reasonable to assume that no language exists in which passive semantics are encoded with zero-morphemes and bare forms while active semantics require morphemic signaling. Rather, active semantics are basic in the cross-linguistic sense, while overtly marking the passive is a path languages may take to varying extents.

As such, the PMDs that Gil describes vastly reduce the complex feature par excellence of Malay, its valence-sensitive derivational morphemes—also recall Adelaar and Prentice's inclusion of the decline of *ter-* and *ber-*—in favor of a simplified rendition of their functions. To this we must add the simplification of the pronominal battery to one which varies lexically only for person while plurality is indicated with a following invariant morpheme, and the use of one morpheme as a general preposition. Adelaar and Prentice also describe additional simplificatory features in eastern PMD varieties, such as substituting schwa for [a] and [e], and frequent loss of final stops and neutralization of nasalization.

### 8.3. *Sociohistory and the PMDs*

Why is spoken Malay usually so different from the standard, and in a simplificatory way? Germane is the fact that only 7 percent of Indonesians speak Malay as a first language (and 45 percent of Malaysians and 15 percent of Singaporeans) (Prentice 1987, 915). Indeed, some Malay specialists assume that the PMDs are the result not of uninterrupted transformation of Malay over time, but interrupted transmission. That is, Riau Indonesian to standard Malay is not French to Latin, but more akin to Haitian Creole to Latin.

For example, in reference to Jakarta Malay, Grijns (1991, 4) concludes that it “developed from Malay as a second language into a Malay-based first language,” noting that in the 1800s under Java’s period of Dutch rule, vast numbers of domestic slaves from various regions and assorted speakers of Chinese varieties were shifting to Malay as first language. Gil (2001, 330–34) notes that even today, one in four Riau Indonesian speakers grew up in homes where at least one parent was not a native Malay speaker, and that “the present-day Riau province was the venue of substantial language contact over much of the last 2000 years,” that “various contact varieties of Malayic must have arisen during this lengthy period,” and that “such contact varieties constitute plausible ancestors for what is now Riau Indonesian.”

But later in the article in question (*ibid.*, 358–67) Gil argues that Riau Indonesian’s simplicity is due not to any form of language reduction, but to three other possible factors. I find all of them problematic.

### 8.3.1. GIVÓN’S PRAGMATIC MODE

For one Gil proposes that Riau’s streamlined nature is due to its status as an oral language, thus used only in what Givón (1979) terms the “pragmatic mode,” characterized by topic-comment structure, short utterance length, minimal embedding, and heavy reliance on context. But most of the world’s languages are used almost exclusively orally. And yet these include languages of awesome complexity in morphology and beyond: that is, all Native American and Australian languages and most Caucasian ones. Thus oral usage alone cannot explain why Riau Indonesian is so telegraphic that, as Gil notes (2001, 325), he has been asked in presentations whether the language was a creole. Givón’s pragmatic mode refers to features such as topic-comment structure, juxtaposition over embedding, and other features, all of which are amply documented in indigenous languages whose grammars with so much inflection according to fine-grained semantic categories, with this then further complicated in interactions with the syntax, that they can leave one marveling that the grammars are even acquirable (e.g., Caucasian or Algonquian languages). Givón’s pragmatic mode does not specify radical elision of morphology.

Gil (2001, 362–63) argues that Riau Indonesian’s simplicity is also an epiphenomenon of the fact that Standard Indonesian is already quite low on inflection and attendant allomorphy. This could be seen as explaining why “pragmatic mode” usage of Malay results in a language reminiscent of a creole while “pragmatic mode” usage in, say, Mohawk results in a strikingly complex language nevertheless. But then Gil notes that Standard Indonesian is, itself, a contact language, and here I would agree. The fact that non-native acquisition of Malay results in so strikingly simplified a language only brings us back to the question as to why the original language was so streamlined to begin with. My argument in this chapter is that this, too, resulted from heavy non-native acquisition.

Finally, we do not see such stark simplicity contrasts between standards and colloquial varieties in other ITs, a relevant example being Smith-Hefner’s (1988) description of the Javanese of Tengger, differing from the standard in minor elisions and collapsings of morphology, but hardly to the extent of Riau Indonesian from the standard.

## 8.3.2. SPRACHBUND FEATURES

As for Gil's argument that Riau Indonesian's nature is due to Sprachbund features typical of Southeast Asia (Matisoff 2001), highly inflected languages like Tagalog, whose grammar he mentions as symptomatic of the tendency for languages in Southeast Asia to elide NP-arguments, are equally typical of the Sprachbund in question. Moreover, this same Sprachbund includes languages of the Tibeto-Burman and Mon-Khmer families which, despite their analyticity, have ergativity and massive lists of numeral classifiers. Our question, then, is why Riau is so naked even beyond its tendency to elide arguments unnecessary to comprehension.

## 8.3.3. NATURAL DEVELOPMENT

Gil (2002) also cites Benjamin (1993) in a thesis saying that the functions of *di-* and *N-* as pragmatic highlighters of patient and actor are the diachronically original situation, and that their conventionalization as syntactic voice markers was a later development in standard Malay.

Intriguing as this idea is, it would seem to founder when examined within the context of Western Malayo-Polynesian as a whole. The account is marginally plausible for *di-* given that it appears to have emerged within Malayic. But *N-* (< *meN-*) is uncontroversially traced back to an affix already conventionalized and syntacticized in Proto-Austronesian.

Benjamin's proposal, then, requires that Malay recast deeply syntacticized inflection *meN-* as a pragmatic marker, and then pedants resyntacticized it on the model of European languages' grammaticalized active/passive distinctions. But this reconstruction casts Malay as an outlier in terms of powerful tendencies in language change.

Gil terms the functions of *di-* and *N-* in Riau Indonesian and like varieties as "semantic," but more conventionally, their optional and discourse-based highlighting behavior would be treated as pragmatic, encoding information structure. As such, Benjamin's account of *meN-* contradicts the strong directional tendency in grammaticalization in which the pragmatic becomes grammatical but the reverse is vanishingly rare (rather than, as Gil has it, exemplifying typical grammaticalization from the "semantic" to the grammatical). That is, Benjamin is proposing that this prefix's development involved an unusual loosening of its usage from the constrained realm of the grammatical into the more optional realm of the pragmatic. Yet the very message of the now vast literature on grammaticalization is that items tend, very strongly, to get caught in a vortex of structural boundedness. And as such, surely to accept Benjamin's account we require that at least a few examples of this rather mysterious inside-out development from grammatical to pragmatic occurred somewhere else in, at least, Western Malayo-Polynesian.

But in his presentation of his pragmatic analysis of *di-* and *N-* in Riau Indonesian, Gil gives no analogous cases in other IT languages despite his career's familiarity with the group. This implies that the proposed development in, at least, *meN-* is unprecedented in IT, and I have encountered no such cases elsewhere in IT in my research. That *meN-* has apparently developed this behavior only in colloquial dialects of Malay suggests that its syntactic use in the standard was indeed the original

one and the pragmatic one a later developmental quirk, caused by an irregular intervention that we would expect to cause such “unraveling.” Only if Malay were considered one of the founding languages of Western Malayo-Polynesian would we be able to save an account of the behavior of *N-* in Riau and other PMDs as an original state. But in reality Malay is a relative latecomer in a group assumed to have begun in the Philippines.

This returns us to asking why a grammatical marker would devolve into the pragmatic only in Malay, and non-native acquisition is a ready explanation. In creoles, for example, categories obligatorily and even redundantly expressed in source languages are very often marked in a smaller range of contexts in the creole, and often only for explicitness. For example, Palenquero Creole Spanish inherits the *ma-* plural affix from Kikongo. But whereas Kikongo marks the plural categorically as does Spanish, Palenquero often omits it when context supplies the inference:

- (38) (a) Kasi to **ma** moná di ayá la baho  
 almost all PL girls of there there below  
 ‘almost all the girls from down there’  
 (b) Ese ea ø mamá puñera-ba re akí.  
 that be.PAST mother boxing-PAST of here  
 ‘Those were the boxing women here.’ (Schwegler 1998, 289)

In cases like pidginization and creolization, “unraveling” of this kind is ordinary. In uninterrupted language change, it is anomalous.

Thus the Benjamin account of *meN-* is highly unlikely. *MeN-* submits more gracefully to the more intuitive account, in which it devolved into the pragmatic secondarily in the PMD Malays as the result of non-native acquisition.

To be sure, *di-* does not trace to Proto-Austronesian or early IT. But then, it submits gracefully to analysis as an analogical development within morphosyntax rather than the pragmatic mode. In the first and second persons, object orientation in Malay is indicated with proclitics, as in:

- (39) Buku ini sudah **ku-**baca.  
 book this PERF 1-read  
 ‘I have read this book.’ (Sneddon 1996, 166)

But in the third person, the clitic is postposed, as *-nya*. The emergence of the rule requiring that *-nya* in this usage occur with preposed *di-* would be due, then, to an impulse to make object-orientation marking match the other two persons in terms of the position of the clitic, with a third-person proclitic derived from the ready source of free pronominal *dia*, and thus

- (40) Narti **di-**tunggu-**nya**.  
 Narti OO-wait-3  
 ‘He is waiting for Narti.’ (ibid., 166)

Such a historical account is in fact a useful explanation for an otherwise peculiar co-occurrence stipulation between *di-* and *-nya* in such cases. Meanwhile, the reconstruction of *di-* as originating as a pragmatic marker has no explanatory benefit—nothing independently suggests that this account is correct; it fills in no explanatory gap; it pro-

vides an answer to no conundrum. And this only brings us back to all of the evidence that the history of Malay has been driven by simplification. If *di-* did, by chance, begin as a pragmatic marker and become syntactic, then this was a distinct exception amid a general process of devolution and streamlining. As such, we require either empirical support or explanatory elegance. Benjamin's account lacks both.

Thus Benjamin's account cannot qualify as counterevidence to a claim that sociohistory transformed Malay at some point.

#### 8.3.4. PARSING HOLES IN THE DATA

A final observation by Gil must also be seen in light of the above arguments. Gil (2001, 360–62) notes that Siak Malay, another local variety, is reduced to a similar extent to Riau Indonesian although there is no especial ethnic heterogeneity among its speakers. To Gil this suggests that Riau Indonesian's simplicity cannot be traced to non-native acquisition. But Occam's razor suggests otherwise.

We will likely never have histories of each Malay-speaking region detailed enough to chart interethnic contact over millennia and their effects upon Malay. Like the abbreviation of Persian, these things happened beyond the purview of writing. But to the extent that most of the regions where PMDs are spoken are ethnically heterogeneous now, or can be documented to have been so in the past, one approach will be to accept that non-native acquisition was responsible for PMDs' simplicity overall.

This is for the same reason that Benjamin's (1993) hypothesis is difficult to accept. If the nature of the PMDs were a typical result of uninterrupted transmission, then we would find several varieties of similar nature derived from other IT languages, or even other Western Malayo-Polynesian languages. But to my knowledge, there are no varieties of, for example, Batak or Buginese that are so much simpler than the standard languages that at first glance they suggest some kind of pidginization. Some Philippines-type languages are more morphologically elaborated than the others, but this variation does not extend to ones with virtually no morphology.

Rather, only Malay presents numerous varieties of this kind — when Malay has the most non-native speakers of any language in Indonesia. Correlation is not cause, but given all of the other evidence we have seen, the chance would seem quite high that in this case, we are dealing with a causal relationship between non-native acquisition and grammatical structure.

#### 8.3.5. A THEORETICAL INCONSISTENCY?

Overall, Gil's perspective on Riau Indonesian and similarly reduced Malays requires that this kind of starkly telegraphic, underspecified homology is merely one pathway that natural languages might take in their diachronic development. The idea, then, is that while some languages might become as complexified as Navajo or Chukchi, that others might take the reverse path and shed massive amounts of grammatical machinery and become as streamlined as Riau Indonesian via uninterrupted development. Gil (2005) states that "the accretion of complexity cannot be construed as an inexorable monotonic process," since apparently "at some stage between Proto-Austronesian and Riau Indonesian, the accretion of complexity must have been reversed."

But even Gil goes on to note that this leaves a question as to “why more languages could not have taken the same path,” and leaves it at “I have no answer to this question.” In this book, I propose an answer: namely, that such a diachrony is simply impossible; as such, I suggest that we assume that Riau Indonesian is a product of interrupted transmission, and that its sociohistory constitutes a clear argument for this. After all, linguists are now well familiar with how and why languages accrete overspecifications, structural elaborations, and irregularity: this is the very heart of the historical linguistic endeavor, especially as including grammaticalization theory as developed over the past few decades.

But it is more difficult to specify in a theoretically coherent fashion just why a language would drift into shedding its machinery on all levels of grammar. Certainly phonetic erosion eternally wears away at a language’s accretions—but just as certainly, grammaticalization and reanalysis at the same time build up new material. Linguists are familiar with cases where this countervailing constructive impulse is retarded to an extreme degree: namely, the emergence of pidgin and creole languages. However, diachronic linguistic theory presents no explanation as to why this would occur without the break in transmission that is diagnostic of pidginization and creolization. Just what would lead speakers of a language to, over the generations, simply toss away morphemes and constructions that encode vibrant aspects of meaning? Note that there is no linguistic terminology referring to such a thing beyond the likes of *pidginization*. This is because there is no known systematic process under which this kind of massive devolution occurs amid ordinary generation-to-generation transmission.

To wit, the idea that languages may opt to either accrete or dissolve is foreign to area specialists worldwide. Bantuists report no Bantu language in central or southern Africa that is mysteriously analytic amid its famously agglutinative sisters—other than Kituba and certain varieties of Lingala, conventionally agreed to be the result of reduction amid non-native acquisition, and most discussed, with little contention, as pidgin or creole varieties. Specialists in languages of the Caucasus document no language that has gone in for undressing while Georgian, Tsez, Ingush, and the rest remain inflectional. There is no isolating Uralic or Australian tongue. There is no Semitic language that has washed away all signs of triconsonantal verb templates—unless we include in that group Nubi *Creole* Arabic. No Sinitic variety eschews tone.

It is in this light that we must assess whether Malay can have “taken it all off” just as a matter of happenstance. Why only this language in this one of countless language families worldwide? And this brings us back to the fact that Malay is, after all, one of the world’s *lingua francas* par excellence. If it is the nature of language change to accrete, then doesn’t a variety like Riau Indonesian necessarily submit to an analysis as deeply impacted by the ravages of second-language acquisition?

## 9. Crucial Issue: Mixture versus Reduction

Given his conclusions and aspects of his wording in his discussion of the role of “contact” in Riau Indonesian’s development, I sense that Gil means that besides pragmatic expression and Sprachbund mimicry (he adduces Benjamin [1993] only

tentatively), language *mixture* was the main factor distinguishing Riau Indonesian from the standard.

This would seem to demonstrate an assumption among many linguists that this book is designed to cut through. That is the guiding idea that when languages come into contact, unless the social situation occasions usage so limited that makeshift pidgins emerge, then the result is the *blending* of grammars rather than their concurrent structural *reduction*.

But despite stressing that Riau has long been used by speakers of many languages, Gil presents no features that Riau Indonesian has presumably borrowed from those languages, as opposed to more general traits supposedly due to general Sprachbund features—which presumably were already locked in much earlier in Proto-Malayic, early IT, or even before. In fact, the spare nature of Riau Indonesian’s grammar is such that we might wonder just what concrete feature it might have incorporated from Minangkabau, Batak, Buginese, and so forth. In the meantime, this book has been devoted to demonstrating that degrees of reduction less extreme than that resulting from pidginization or creolization can occur as the result of widespread non-native acquisition.

### 9.1. *Mixture Alone Undergenerates the Data*

Let us allow, in any case, that Riau Indonesian is “mixed” not in that it has inherited specific features from other languages, but has simply taken on the underspecified, highly context-dependent nature of surrounding languages’ general game plan. The problem is that this kind of homology also lends itself to being traced to simplification rather than mixture. So—how can we decide which analysis is right?

In that light, it must be recalled that languages can blend without the result being simplified in any fashion that would even occasion attention. A “mixed” language like Media Lengua combines Spanish lexicon with a faithful reproduction of Quechua morphology and syntax (Muysken 1997); Media Lengua grammar is as complex as Quechua’s. Australian languages have exchanged morphology so rampantly that family trees are almost impossible to reconstruct (Heath 1978; Dixon 2001), with no effect upon their general level of morphological elaboration (amid which there are no reported “Malays” mysteriously shedding all but a hint of their ancestors’ inflections). In terms of analytic languages of Southeast Asia, useful are cases like the southern Chinese, which have been deeply influenced by languages of other families in the area—but the result is languages which, while telegraphically underspecified in the fashion that Gil calls attention to as a Sprachbund feature, remain massively complex on other levels such as numeral classifier proliferation, number of negator morphemes, fine gradation of the aspectual space, and so on.

Thus to parse Riau Indonesian as simply a hybrid, we must identify which type of hybrid it is. Is it a “mixed language” like Media Lengua? Or has it borrowed massively but retained its level of complexity, like Romanian—or southern Chinese? Or does it fall somewhere on the creole continuum, in which simplification is as prevalent as borrowing, or, in some cases—a crucial potential comparison in this case—simplification is all that happened, with little or no identifiable substrate transfer (as is the case in Hawaiian Creole English)?



Our question re the PMD Malays, then, is which situation it more closely parallels in terms of reduction, and it would seem to be the creole one. Riau Indonesian, for example, has numeral classifiers—but to a small and fitful extent. It is typical of the Southeast Asian Sprachbund for roots to be flexible between adaptation as nouns or verbs, but Riau takes this to an extreme, with its cross-linguistically striking degree of zero-derivation. That is, the main course in Riau Indonesian's development was simplification, not mixture.

### 9.2. *Malay Mixed but Simplified: Baba Malay*

Thus like in creoles, even to the extent that other PMDs bear more robust signs of mixture with other grammars and of more recent origin than mere Sprachbund tendencies toward underspecification, these PMDs remain at the same time distinctly simplified in comparison to those grammars as well as Malay's.

It is tempting to deflect the reduction as a background factor, in favor of stressing the mixture as an example of the grammatical hybridity that so often interests various analysts. But this approach misses stark qualitative differences in ratios of mixture to reduction. This becomes evident in clear examples of Malay mixing with other grammars.

On the one hand, there are the Chamic languages of Vietnam, which resulted when an Austronesian language came into contact with Mon-Khmer ones. Lexical reconstruction shows that the Austronesian language in question was a relative of Malay, such that Thurgood (1999) reconstructs a Malayo-Chamic ancestor to Malayic and Chamic. Yet with their multilevel systems of tone or register and other inheritances from Mon-Khmer grammatical structure, the Chamic languages give no more impression of simplification compared to other grammars than Mon-Khmer—or most Austronesian—languages.

It is to cases like this that we must compare the contrasting case of PMD Malays. Crucial are PMD varieties with—unlike Riau Indonesian and other PMD Malays Gil describes—clear inheritances from recent contact with specific other languages, but still saliently less complex than either standard Malay or the source languages.

Baba Malay of Malaysia and Singapore is typical here. It was developed by Hokkien (Min) Chinese speakers, and, as a result has some Hokkien-derived features such as phonemic nasal vowels, some pronouns, and a comparative construction, and, possibly, extensions of the use of the *ada* existential, the *punya* possessive, and the use of *kena* “give” for the passive (Ansaldo and Matthews 1999).

But if Baba Malay were simply a mixture of Bazaar Malay and Southern Min with reduction insignificant, then we would expect it to display, if not seven or eight lexical tones like Min, then at least three or four (Western Cham, for example, as an Austronesian-Mon-Khmer hybrid, has five). We might also expect a healthy system of grammaticalized serial verb constructions, multiple semantically specific negators, a reflection of Min's multiple complementizers, and/or other Min features we encountered in chapter 4.

That is, if Baba Malay were simply a blend of Bazaar Malay and Southern Min with structural reduction limited to that in the long-ago emergence of Bazaar Malay, then there are two possible expectations we might have:



1. Baba Malay would give a preliminary impression of being a Chinese language with Malay words. Analogies: Sinitic varieties infused with Altaic influence like Hezhou and Tangwang; Chamic varieties whose very classification as Austronesian was once an open question.

2. Less extremely, we might expect that Baba Malay, while still identifiably a variety of Malay, would be replete with phonological and syntactic features distinctly unlike Malay or IT ones, and clearly derived from Chinese, obviously enough that there would be no need for analysts like Ansaldo and Matthews to carefully outline an argument for the point. Analogies: Slavic inheritances in Romanian and Yiddish that would stand out as odd compared to the languages' sisters even if Slavic languages no longer existed, or Dravidian features in Indo-Aryan languages, unlike anything encountered anywhere else in Indo-European.

But neither of these alternatives is remotely in evidence. The Hokkien features in Baba Malay are scattered (Ansaldo and Matthews present a mere seven structural features) and shallow, tincturing assorted fragments of grammar but virtually never occasioning constructions utterly foreign to standard or other PMD Malays. Thus where a PMD variety like Riau Indonesian uses *kena* as a passive marker without an expressed agent:

- (41) Mister      kena      tembak.  
white.man undergo shoot  
'The white man got shot.' (Gil, forthcoming)

Ansaldo and Matthews present a sentence in which Baba Malay uses *kena* with an expressed agent, modeled according to them on a similar behavior of Hokkien *hō*:

- (42) Nkoh    gua kena orang tipu.  
brother my "by" people cheat  
'My brother was cheated.'

But this is the only sentence of this kind that they give, from what they acknowledge is a limited, written corpus. And in it, *kena* is potentially interpretable as a full verb after which *orang* is an unmarked argument, given how chary PMD Malays are of marking grammatical role. In this light we might note that the translation of the sentence is somewhat terse, that the fuller one Ansaldo and Matthews give, "My brother was cheated by people," is somewhat infelicitous in terms of naturalness, and that the connotation may possibly have been one with an overtly referential NP adjunct "My brother got cheated by some people," in which case we would have a usage of *kena* similar to the Riau Indonesian one.

More divergent from standard or PMD Malay, Baba Malay can extend possessive *punya* to attributive constructions, as in:

- (43) betol **punia** hap  
good POSS match  
'a good match!'

It can also extend along the lines of Hokkien:

- (44) khăq tuā é      cābò  
 more old PART woman  
 'An older woman.'

This is indeed rather quirky compared to standard Malay—but hardly remarkable given the divergence from standard Malay typical in PMDs Malays, for example, *sama* as a general preposition, reduced forms of *ini* and *itu* as determiners. To wit: at the end of the day, the features that Ansaldi and Matthews (1999) trace to Hokkien—seven structural hiccups—would hardly seem anomalous in a Malay variety if Baba Malay's creation by Hokkien speakers were lost to history. In that case they would seem, instead, mere variations upon a familiar pattern. Recall, after all, that there are varieties of Malay that vary far beyond this degree, such as Kerinci Malay with its nouns whose shapes vary according to position reminiscently of consonant gradation in Baltic-Finnic. No careful argument is necessary that this Kerinci development represents a sharp departure from standard Malay or even nonstandard Malays. In the same way, no analyst could be presented with Chamic and let it pass as just business as usual in Malayic.

Despite, then, its slight and subtle distortions from Hokkien, Baba Malay has very much the profile of a PMD Malay. It contrasts with standard Malay most in its reduction, not in its scattered parallels to Hokkien.

### 9.3. PMD Malay: The Big Picture

Some prefer to analyze the PMD prototype itself beyond just Baba Malay as influenced by southern Chinese languages, given that the use of *orang* to pluralize pronouns, the *give* passive/causative, and the *punya* possessive linker all have equivalents in Chinese grammar (e.g., Adelaar and Prentice 1996, 674). It is "influenced," perhaps. But we must keep in mind how much more deeply Chinese influence can penetrate a grammar. Examples include, again, Chamic, or even Vietnamese, thought to have inherited its phonotactics and tonal richness from Chinese. Why aren't the PMD Malays tonal like Chamic? Why aren't PMD Malay grammars so deeply Chinese in typology that they confuse taxonomists?

If PMD features are modeled on Chinese, it is to a distinctly moderate extent—of the sort that requires smoking out by the careful analyst. But this is one of many possible outcomes of a synthetic, toneless grammar's contact with analytic, tonal languages such as Sinitic and Mon-Khmer ones. Thus we can allow that at least some PMD Malays bear hallmarks of contact with other grammars such as Chinese, while also realizing that the signs of that contact do not, properly, refute an account stipulating that the PMD Malays were crucially shaped by non-native acquisition of Malay. That is, PMD Malay and/or Bazaar Malay is not simply Malay mixed with Chinese; it is Malay whose Chinese-speaking learners affected it slightly in minor structural transfer, but mostly in reducing its complexity, as is typical of second-language learners.

Final demonstration: the following two sentences demonstrate Malay mixing with other languages as compared with standard Malay. The first is the colloquial Malay

of the island of Roti, Basa Kupang, compared to the same sentence in standard Malay (Fox 1982, 312, rearranged):

- (45) Basa Kupang    Beta su       pi       ma lu       sonde ada.  
          Standard Malay   Saya sudah pergi tapi kamu tidak ada.  
                                  I     PERF go       but you   NEG be  
                                  ‘I went but you weren’t there.’

The mixture here is in the form of borrowings: *lu* from Hokkien Min; *sonde* is from Dutch *zonder* (Kumudawati and Setjadiningrat 2001, 78); *ma* is likely from Dutch *maar* reinforced by a Rotinese conjunction of similar shape. *Su* and *pi* are shortened forms of the standard ones. But the sentences are structurally identical, neither Min nor Rotinese grammar playing any part. The perfective marker is not postposed as in Min; Rotinese’s subject-marking inflectional prefixes are nowhere in evidence.

Contrast this with a sentence in Acehnese, the result of a language closely related to Malay mixing with Mon-Khmer:

- (46) Na si-droe-ureueng-ladang geu-jak lam-uteuen geu-jak koh kayee.  
          be one-CLAS-person-farm 3S-go in-forest 3S-go cut wood  
          ‘There was a farmer who went into the forest to cut wood.’ (Durie 1985, 192)

The phonetic inventory is unlike that in any IT variety. Orthographically, *eu* represents unrounded back vowel [u]; *eue* represents a diphthong with a schwa offglide ([uə]); *oe* represents [uə]. These are clear parallels to typical sounds in Mon-Khmer languages. The *droe* classifier comes from the use of this word as “self” (cf. Malay *diri*), but Malay’s cognate is not used in this function. The syntax departs from Malay as well, with its obligatory subject-marking clitics and absence of voice-marking prefixes (*di-* and *meN-* have no cognates or equivalents in Acehnese).

Here, then, is a Malay-related language in which mixture has created a language sharply different from any form of Malay itself. This leads to the question as to why, if the PMD Malays differ from standard Malay only in being “mixed,” the mixture in them is so consistently moderate and shallow.

Of course, there are exceptions, such as the Macassarese Malay that intertwines Macassarese grammar with Malay lexicon (Adelaar and Prentice 1996, 682). But this only leads to the question as to why most or even all of the colloquial Malays are not mixed to this extent, if mere mixture were the main factor involved in their emergence.

## 10. What Happened to Malay?

Given the observations of the previous sections, I believe that it is warranted to assume that Malay, in both its standard and colloquial forms, is the product of incomplete transmission of Proto-Western Malayo-Polynesian grammar. Given Malay’s longtime status as a regional lingua franca, it is most economical to suppose that this status was the causal factor.

### 10.1. *Malay through the Ages*

There is controversy over whether Malay emerged in Sumatra or Borneo, with consensus converging upon the latter (e.g., Adelaar 1992, 207). However, eventually its speakers occupied both sides of the Strait of Malacca, between Sumatra and Malaysia. Trade between Arabs and Indians in the west and the Chinese in the east required passing through this strait, and the monsoon season required travelers to pause for months in the region, ensuring even more contact with Malay.

The first inscriptions in Malay date to the 680s, under the Srivijaya Empire that covered most of present-day Malaysia and Western Indonesia and had its capital in Palembang in southern Sumatra. While this is the first concrete evidence of Malay's use as a lingua franca, the presence of Malay words in Malagasy suggests that Malay was widespread at least centuries before this, as Malagasy traces to languages of south-eastern Borneo, whose speakers likely migrated to Madagascar around the 300s. In the centuries after the Srivijaya inscriptions there are inscriptions from Malaysia, Java, and other parts of Sumatra.

The regional center of power was to change location, first to Malacca in Malaysia from the late 1300s to the early 1600s, and then to an area comprising the southern tip of Malaysia (Johore) and nearby island groups (Riau and Lingga). These centers produced literatures in Classical Malay, which served as a literary variety and was, predictably, conservative and only approximately reflected casual speech. Meanwhile, the colloquial Malays were proliferating beneath the radar screen of writing (historical data from Prentice 1987; and Adelaar and Prentice 1996).

### 10.2. *Malay Sociohistory and Bazaar Malay*

It would seem, then, that Malay underwent two "passes" of reduction. The literary Malay of the courts and literature, while slightly more complex than modern Malay, was still unusually reduced compared to earlier IT languages. Presumably this resulted from the language's widespread and long-term use by non-native speakers simplifying it before it was committed to the written medium.

Then, the colloquial Malays were reduced even further, as the result of continued heavy impact from non-native acquisition. The conventional assumption (*pace* Gil and some others) is that these varieties are outgrowths of the Bazaar Malay pidgin. However, this is explicitly indicated neither in documentation nor in contemporary observations, neither of which offer remotely enough detail for any conclusions on this score. And in actuality, the fact that both Bazaar Malay and the colloquial Malays exhibit highly reduced features does not entail that varieties like Riau Indonesian, Baba Malay, and others began as Bazaar Malay.

It is equally possible that these varieties emerged with the same intermediate degree of reduction as they have now, just as the new Arabics need not have arisen from an established pidgin. It is clear to most observers that synchronically, the results of language simplification exhibit themselves to degrees. However, it is less often clear that the process itself occurs to a similarly gradient degree, rather than necessarily beginning at the pidginization level and then "bouncing back" from there. That is, when a variety of a language is structurally reduced in comparison to the standard to a degree

between pidginization and full transmission, there are no grounds for an automatic assumption that the variety must have begun as a pidgin and elaborified later.

To be sure, there are examples where this is the case, as in Tok Pisin and its sister creoles, and as argued for most plantation creoles in McWhorter and Parkvall (2002). But just as often, reduction occurred from the outset to a moderate but perceptible degree, as is clear from the history of Réunionnais (McWhorter 2000b, 187–90) and African American Vernacular English (McWhorter 2001).

Historical evidence will likely never tell us the extent to which Bazaar Malay played a role in the development of colloquial Malays in the dozens of relevant regions. However, to the extent that Bazaar Malay played a role, its roots in non-native acquisition are incontestable. Meanwhile, however, one implication of my argument in this chapter is that even if Bazaar Malay had never existed, Occam's razor would require us to trace the colloquial Malays to the same factor. In any case, likely in each or most locations Bazaar Malay existed in a complex interaction with varying degrees of acquisition of Malay. But the crucial observation is that the contrast in grammatical elaboration between the PMD Malay dialects and all of the other IT languages is vanishingly unlikely to have arisen as a mere result of uninterrupted internal change, even mediated by "pragmatic mode" usage.

### 10.3. *Grounds for the Hypothesis*

We must assume, then, that both standard and PMD Malay are the results of grammatical simplification caused by incomplete acquisition. There are two discoveries that would refute this hypothesis.

#### 10.3.1. REFUTATION ONE: RADICAL SIMPLIFICATION ELSEWHERE IN IT

One would be IT languages which had undergone no significant non-native acquisition over their histories, which were nevertheless similarly reduced compared to Ross's early IT reconstruction and in other features such as those in living IT languages such as the ones of Sulawesi. Surely if the transition from Ross's early IT reconstruction to standard Malay is simply business as usual in the evolution of IT languages, then we would expect that not just one, but at least several groups of isolated IT speakers would speak languages similarly abbreviated compared to a *Tukang Besi* or a *Batak* variety.

The same applies to varieties like Riau Indonesian. If this homology can result from uninterrupted transmission of a grammar, then we would expect at least a few other IT languages of similar grammatical profile. To my knowledge, at present there has been discovered not even a single one such variety.

#### 10.3.2. REFUTATION TWO: RADICALLY CONSERVATIVE MALAYS

A second refutation would be several nonstandard Malay varieties spoken in regions with a long tradition of interethnic mixture that have ample overspecifications,

structural elaborations, and/or irregularities absent in the standard. I refer to a hypothetical Malay in which, perhaps, there are obligatory subject-marking prefixes, a three-way distinction in demonstratives, a good dozen numeral classifiers in regular use, and imperative affixes. Varieties like this would vastly weaken the link I propose between non-native acquisition and the contrast between standard and colloquial Malay, and suggest that the difference between Adelaar and Prentice's "PMDs" and these radically conservative dialects was a matter of a mere roll of the dice.

I highly suspect, with all due respect for the difficulty of fieldwork in many of the relevant locations, that if an array of dialects of this kind existed, then we would know by now.

However, I must also stipulate that if just a single dialect of this kind were discovered, then I would consider its impact on my thesis minimal. The very nature of the scientific enterprise is such that we cannot allow the fluke to dissolve a thesis couched in an attempt to explain overriding tendencies. Short of especially comprehensive sociohistorical documentation on the region, there are any number of possible scenarios that would allow the happenstance preservation of such a variety, such that the overwhelming simplificatory tendency in dozens of regions elsewhere would still stand as a norm, and as such, demand explanation.

Moreover, if such a variety were found in an isolated area in Borneo—notoriously understudied linguistically—then this could well be seen as long-sought proof that Malay originated there. This variety would qualify as a valuable survival of Malay unaffected by non-native acquisition—as a Min compared to Mandarin or an Icelandic compared to English. However, vast predominance elsewhere of abbreviated renditions of standard Malay grammar would still submit to an account based in non-native acquisition.

Thus a single wild card will likely not qualify as a refutation. We seek a number of Malay dialects counterbalancing the PMD ones, numerous enough to confirm that the degree and type of interethnic mixture that Malay has undergone was not sufficient to abbreviate its grammar to a regular extent. It would be unrealistic to require that the number of such varieties equal the number of PMD varieties now known, especially given the minimal probability that so many could exist and be as yet undiscovered. But it is realistic to suppose that a good three or more such varieties would have to be uncovered—optimally in geographically separate locations to rule out contact effects between them—before my thesis would be considered refuted.<sup>6</sup>

## 11. Loose End: Malayo-Javanic

In this chapter I have drawn particular attention to the fact that the virtual eclipse in Malay of allomorphic paradigms of bound morphemes marking certain distinctions is a developmental anomaly. The Philippines-type languages have paradigms of affixes linked to focus on constituent class; in most IT languages this disappears, but is replaced by allomorphic affixes marking concord with subjects and/or objects. In Malay the only reflection of this is number-neutral clitics indexed to person, but used

as arguments rather than concordial markers, and even then replaceable with free pronouns (section 4.7).

A general observation in this monograph is that radical loss of inflectional morphology across a grammar is not a “normal” development, and instead almost always signals an interruption in transmission. But in the light of our IT data, a question remains, then, about Malayic varieties Minangkabau and Iban and the “Malayo-Javanic” languages of Java, Javanese, Madurese, and Sundanese. Like Malay they lack the concordial affixes so typical elsewhere in IT, such that their grammars parallel Malay’s to the extent reminiscent of the Romance languages. This might be seen as an indication that this absence in Malay is indeed unexceptionable — just one of many paths an IT language might take.

### 11.1. *The Complexity of Malayo-Javanic beyond Malay*

In addressing this, it must first be recalled that the three languages of Java, in particular, nevertheless surpass Malay saliently in overspecification, structural elaboration, and irregularity. All three retain a productive infix (in Sundanese, verbs take *-ar-* in the plural [Müller-Gotama 2001, 20]), and all three have a three-way demonstrative gradation. Javanese has imperative morphology and some suppletive plural nouns; Sundanese has the typical Malayo-Polynesian battery of numeral classifiers; Madurese has aspect-marking inflections; and so on. Also relevant is Javanese’s famous three levels of honorific marking in its lexicon, according to “high” (*krama*), “middle” (*madya*), and “low” (*ngoko*) (Geertz 1968, 284):

#### (47) Honorific Levels in Javanese

<i>krama</i>	menapa	pandjenengan	baḍé	ḍahar	sekul	kalijan	kaspé	samenika
<i>madya</i>	napa	sampéjan	adjeng	neḍa	sekul	lan	kaspé	saniki
<i>ngoko</i>	apa	kowé	arep	mangan	sega	lan	kaspé	saiki
	are	you	going	to eat	rice	and	cassava	now

‘Are you going to eat rice and cassava now?’

While usually treated as a sociolinguistic phenomenon, in the structural sense this is a significant overspecification, overtly marking social level to a discrete and obligatory extent that most languages do not, just as other languages may overspecify distinctions such as referentiality, tense, evidentiality, and so forth. (That Madurese and Sundanese, as well as nearby Balinese and, to an extent, Sasak have inherited this trait from Javanese [in borrowing the “high” lexicon directly] qualifies as a borrowing rather than as independent overspecification.)

### 11.2. *Simplicity in Malayo-Javanic beyond Malay*

But the fact remains that the languages of Java lack the concordial bound morphemes found so consistently across most of IT. Moreover, documentation of earlier stages of Javanese, while capturing the language after this particular loss had already occurred, suggests that the modern state of the language developed from a more overspecified language more typical of the IT profile. Old Javanese (of the 800s to 1400s) retained two productive infixes now lost, *-um-* with a wide range of meanings

(*santwa* “respect,” *sumantwa* “treat with respect” [Teselkin 1972, 43]) and *-in-* for the passive (*pangan* “eat,” *pinangan* “to be eaten” [ibid., 44]). Old Javanese retained a definiteness marker *ang (ng)* (ibid., 51–52) as an echo of the focus-marking function of this morpheme in Philippines-type languages, and similarly also marked actors when not used as subjects:

- (48) Ya        ta        cinakrākēn    **de**        bhaṭāra Wiṣṇu    ring    daitya.  
 EMPH EMPH hit.TRANS ACTOR Lord    Vishnu OBL discus  
 ‘Lord Vishnu hit (the giant) with the discus.’ (ibid., 59)

Old Javanese was predominantly VSO like Philippines languages and most IT ones, but Poedjosoedarmo (2002, 325) analyzes the eclipse of grammatical role markers like *de* as having forced the default SVO order of modern Javanese, paralleling a similar development in Malay but unlike IT languages beyond the “Malayo-Javanic” complex, where VSO, a hallmark of Western Malayo-Polynesian, dominates.

The question is why Javanese lost these features when countless Philippines-type languages have retained, for example, ample inflectional morphology, complete with allomorphic paradigms, for millennia? What happened to Javanese?

In assessing the exact phenomenon we must account for, we observe that most of the abbreviation of the IT legacy in Javanese, as well as Sundanese, Madurese, Minangkabau, and Iban, is inflectional. Here we have an example of the fact that inflection is the “front line” of simplification in language contact. Afrikaans has lost most of Proto-Germanic’s inflection, but otherwise retains almost as much of the Proto-Germanic legacy as other Germanic languages, while English lost not only the inflection but also a great deal more (see chapter 2). Meanwhile, there is no Germanic language vastly decomplexified compared to Proto-Germanic but retaining most of its inflectional paradigms. There are creoles that retain much of their lexifier’s derivational morphology but lack its inflections, but none that retain the inflections but lack the derivation. Thus our specific question is why the other Malayic languages and the ones of Java abbreviated the IT legacy even to the preliminary extent of eschewing concordial morphemes.

Here, it is crucial that we see such loss in IT languages other than Malay only in languages spoken closely to it and with long histories of intense contact with it.

### 11.2.1. SIMPLIFICATION IN PROTO-MALAYO-JAVANIC?

In searching for the locus of the transformation, we might look to the ancestor of all of these languages, supposing that Proto-Malayo-Javanic was a uniquely analytic language amid IT. We might even surmise that this development was a “natural” one. Certainly, subfamilies develop quirks, such that we might analogize the absence of concordial inflection in Malayo-Javanic to Grimm’s Law in Germanic. But there are problems here.

First, the analogy is not graceful: while Grimm’s Law entailed a transformation of phonetic materials rather than an abbreviation of them, the Malayo-Javanic languages simply eliminated a feature—paradigms of allomorphs—ordinary and long lived in the hundreds of other IT languages. Second, while Grimm’s Law is a notoriously peculiar quirk cross-linguistically, the anomalous and rapid shedding of in-



flection has analogies in processes like pidginization and creolization, forcing us to consider that externally induced processes of this kind may have been responsible for the Malayo-Javanic trait. In this light it is also relevant that it has been proposed that Grimm's Law resulted from non-native acquisition of some kind (e.g., Vennemann's works arguing for an encounter between Germanic and Semitic, e.g., 2000).

In this light, we might return to a contact explanation and surmise that at the Proto-Malay-Javanic stage there was some historically undocumented interruption in transmission. But since this would have happened before writing provided us with any confirmation linguistic or historical, this explanation would remain permanently speculative. Besides, the losses that we do have documentation for between Old and Modern Javanese suggest that something irregular happened to Javanese much later than Proto-Malay-Javanic.

All of which returns us to the peculiarity of this chariness of concordial markers compared to Western Malayo-Polynesian—and to Austronesian in general. The only IT languages that parallel Malay to any extent in this regard are spoken near it. This suggests that only one language, Malay, took the crucial detour, and that languages spoken nearby were merely impacted from its influence.

But this returns us to the question as to why Malay itself took this detour at all. To suppose that this was a happenstance requires comfort with assuming that the emergence of this homology in just one of thirty-odd groups within Western Malayo-Polynesian was simply one roll of the dice—when, recall, Malay has been a major lingua franca for most (if not all) of its very existence, while all but a few of the IT languages are minority ones spoken by a few thousand.

### 11.2.2. AREAL INFLUENCE

Looking at a map, then, we see inflection falling away solely around the homeland of the IT language most widely spoken nonnatively for millennia, and furthermore to degrees (Javanese retains imperative marking, Madurese retains aspect marking, etc.). The most economical account would be an areal feature, due to contact with the one language whose structure was deeply affected by non-native transmission.

Written history is largely silent as to the particulars of the encounters between Malay and the languages of Java. But lexical borrowings from Malay are rife in them, permeating their lexicons far beyond the level of marginal "cultural" items or any restriction to "high" registers as in the case of the impact of French upon English. According to Thomason and Kaufman's (1988, 74–76) landmark tabulation of degrees of language contact and their effects, this core lexical influence falls at Level Two on their borrowing scale—crucially, at which some concurrent syntactic influence occurs as well. That is, the depth of the penetration of Malay lexicon in the Javanese languages is of a kind that typically occasions grammatical influence as well.

Thomason and Kaufman focus on morphosyntactic *borrowings* rather than *elisions*, noting the difficulty of identifying simplification as the guiding process in instances of language contact that occurred before written history (*ibid.*, 114–15). But

this book is an attempt to fill in this gap, and as such there is nothing in the studies by Thomason and Kaufman or other authors that belies the possibility that syntactic borrowing might involve not only present but also absent features. To wit: there would appear to be no principled reason why areal contact, even when not creating pidgins or creoles, would not result in speakers eliding grammatical features familiar in their native languages that are absent in the contact language.

And there are ample examples in Sprachbund cases. In the Balkan Sprachbund, in one of many examples worldwide of phonological loss in areal influence, all member languages lost length and nasality modeled on Greek; all also lost the morphological infinitive (Schaller 1975). In another case, Australia offers an example of areal contact eliding head-marking concordial affixes analogously to what I propose happened in Malayo-Javanic. The southern dialect of Baagandji of New South Wales has subject clitics indexed to person and number:

- (49) ŋiinga-ŋgu-aba  
sit-PERFECT-1S  
'I sat.'

But Gurnu Baagandji of the north, spoken in a Sprachbund with contiguous languages such as Wangkumara, Badjiri, and Muruwarri that lack bound pronominal forms, leaves the verb bare and has free pronouns that have taken on tense marking:

- (50) W-adhu gaandi barlubarlu.  
PAST-3S carry small.children  
'It was him that carried the small children.' (Dixon 2001, 78)

Elaborate paradigms of noun-class marking prefixes are typical of Niger-Congo, famous in Bantu to the south but also robust in West Atlantic languages northward like Wolof and Fula. These prefixes are fossilized or absent only in a subset of Niger-Congo languages spoken between Senegal and the Bantu region. Specifically, this absence of noun-class marking is found on the border of the family's area that is contiguous to Chadic languages that lack such marking, the best-known examples including Mande and Kwa languages and close relatives of Nigeria like Yoruba and Igbo (Dimmendaal 2001, 378–81).

As such, the concordial loss in Malayo-Javanic, in which the languages otherwise parallel other IT languages in overspecification and structural elaboration, can be seen as a Sprachbund effect. We might propose that Javanese, Madurese, Sundanese, Minangkabau, and Iban are members with Malay of a Sprachbund in which language contact was intense enough that these languages lost concordial marking modeled on Malay, which they maintained at an earlier stage, as Gurnu Baagandji and many Niger-Congo languages lost concordial affixes under influence from surrounding languages. Recall that inflections are on the front lines when it comes to loss amid non-native acquisition, as demonstrated by Afrikaans compared to the rest of Germanic, and by creoles, among which many retain derivation but not inflection, although none retain inflection but not derivation.

There exists, to be sure, no series of documents that demonstrate earlier stages of these languages losing the concord. However, neither is there documentation

of the evolution of the languages of the Balkan Sprachbund, Baagandji, Mande, Kwa, Yoruba, or Igbo neatly recording these languages drifting toward their present-day structural likenesses with nearby ones of other groups. These developments occurred in languages unwritten at the relevant time, as most are even today. Rather, the uncanny structural clusterings among these languages today is considered scientifically significant demonstration that their likeness is not accidental.

I propose that the likeness in concordial anomaly between Malay and the Malayo-Javanic languages today is similarly indicative of a contact-induced relationship. My reason for this is, again, that no other of the three-hundred-plus IT languages has developed the way Malay and the Malayo-Javanic languages have. If Iban, Minangkabau, and the languages of Java simply shed concord by chance, then where are the IT languages spoken far from Malay that are similarly analytic in this regard? And given this, is it really insignificant that the only IT languages of this typology are those spoken closest to Malay?

## 12. Analogous Cases: Central Malayo-Polynesian

The link between Malay's simplicity and non-native acquisition within Western Malayo-Polynesian is supported by similar correlations in some languages in the smaller and lesser-known Central Malayo-Polynesian subfamily.

### 12.1. *Tetun*

For example, there is an Austronesian language where there is a living complexity contrast between a conservative indigenous rendition and a strikingly streamlined lingua franca one, analogous both to standard Malay versus PMD Malays and to Proto-Malayic versus standard Malay.

The case is Tetun of East Timor, of the Central Malayo-Polynesian subfamily.<sup>7</sup> Two varieties are spoken as first languages by small communities: Tetun Terik in a southern coastal region and Tetun Belu on the border between East and West Timor. But Tetun Dili is a lingua franca of the island, spoken by two-thirds of the population, whose native languages are eleven other Central Malayo-Polynesian languages and four distantly related to Papuan ones.

Grammars exist for both Tetun Terik (Van Klinken 1999) and Tetun Dili (e.g., Williams-Van Klinken, Hajek, and Nordlinger 2002). Tetun Dili differs from Tetun Terik partly in predictable effects of contact with the colonial language Portuguese, including lexical and phonemic borrowings. Predictably, many comparisons of the two imply that the main difference between the two varieties is in degree of mixture with Portuguese. But this is only because the assessment of differences in grammatical complexity is as yet only fitfully attended to in language contact studies beyond those of pidgins and creoles.

Upon examination, Tetun Dili (henceforth TD) differs from Tetun Terik (henceforth TT) not just in alternative structural choices but also in a saliently simpler grammar.

## 12.1.1.1. TETUN TERIK VERSUS TETUN DILI

Tetun Terik has three verbal affixes: *ba-* for causative, *hak-* as an intransitivizer, and *mak-* for marking actorhood (*daka* “guard,” *ema mak-daka-n* “person who guards”). In TD, *mak-* is absent, and the *hak-* cognate is present but less obligatory or productive than in TT.

TT uses six numeral classifiers. In TD nouns usually occur without classifiers, and only four are in use at all. TT has a definite article, although its occurrence varies geographically; TD leaves definiteness to context.

In TT, possession entails enclitics on the noun whose choice is determined by constraints of alienability and other factors: *ó fé-r* or *ó fé-n* “your wife,” but *sa ata-r*, \**sa ata-n* “their slave.” Tetun Dili has lost the elaborate conditioning of these clitics, instead using a free morpheme *nia* for possession (*ahau nia liman* “my hand”), with fossilized *-n* as the only remnant of the TT situation (*uma João nian* “John’s house”).

Tetun Terik has three functionally distinct copulas; TD retains just one (*iha*). Tetun Terik has an equative negator and a predicate one:

- (51) (a) *Buat e’e Bei Beur ha’i!*  
           thing this Mr. deceive NEG  
           ‘This thing isn’t Mr. Trickster!’
- (b) *la bele la’o*  
           NEG can walk  
           ‘cannot walk’

Tetun Dili retains just *la* across the board.

Most famously, TT has subject-marking concord prefixes. Their appearance is complexified by morphophonemic rules. For example, the prefixes replace initial /h/, such that *k-sai* “I exit” but with *há* “to eat”, *ká* “I eat” rather than \**k-há*. When the *h*-initial verb is trisyllabic and begins with the *hak-* prefix, the prefix’s /k/ is optionally elided in the first-person singular: *hakdiuk* “play” > *k-akdiuk* or *k-adiuk*.

Tetun Dili, however, in its structural difference from TT most remarked upon, lacks subject-marking prefixes entirely.

## 12.1.1.2. MERE “DRIFT”?

Yet we cannot suppose that rather than TD losing the prefixes, the prefixes were an independent emergence in TT. For one, cognate prefixes are typical of Tetun’s relatives. Besides, TD has verbs that occur only in the third person that begin with *n* as opposed to their cognates in TT beginning with other consonants; since *n* is the third-person singular prefix in TT, this shows that at an earlier stage TD indeed had the prefixes. The three Tetun dialects even show evidence of a diachronic cline of erosion of the prefixes: they occur with all verbs in Tetun Belu, only with verbs beginning with a vowel or *h* in TT, and then are entirely absent in TD (Hull 2001, 152–53). The question is why the elision went so much further in TD than in TT.

Besides, the overall complexity contrast between TD and TT is mirrored by a similar one between TD and related languages. Comparisons with Rotinese, of Tetun’s Fabronic subbranch of its Timoric group within Central Malayo-Polynesian (data from Jonker 1915) and Kambera of the Bima-Sumba group (spoken on Sumba) (data from Klamer 1998) illustrate this.

Both languages have derivational verb prefixes similar to those in TT. Rotinese has eight numeral classifiers, and Kambera six. Both languages have a definite article. Possession in Rotinese involves enclitics conditioned by constituent class (Jonker 1915, 270–74). In Kambera, instead of equative and locative copulas, nonverbal predicates are marked with accusative clitics, marking nonverbal predicates for case in a fashion overspecificational compared to most languages including TT:

- (52) Tau   mini-ya.  
person male-3S.ACC  
‘He is a man.’ (Klamer 1998: 162)

Kambera also marks subjects with the accusative in impersonal constructions, thus rendering an absolutive distinction in a language otherwise nominative/accusative:

- (53) Jàka   nda   nyumu, meti-ya.  
CONJ NEG you   die-3S.ACC  
‘Without you, we would die.’ (one would have died) (ibid. 161)
- (54) Nggàra mài-ya-i                                   nú?  
what   come-3S.ACC-ITER there  
‘What can I do for you?’ (ibid. 167)

Then both Rotinese and Kambera have the subject prefixes, including morphophonemic complications and irregularities. Examples of this are also found in other Central Malayo-Polynesian languages. In Sika of the Flores-Lembata group (spoken in Eastern Flores), morphophonemic processes after the pronominal morphemes create incipient conjugational paradigms (data from Lewis and Grimes 1995, 605–6):

TABLE 8.10. Verb Paradigms in Sika

	<i>bano</i> “go”	<i>gita</i> “see”	<i>raʔit</i> “sew”
1S	pano	ʔita	ʔraʔit
2S	bano	gita	raʔit
3S	bano	gita	raʔit
1P (exclusive)	bano	gita	raʔit
1P (inclusive)	pano	ʔita	raʔit
2P	bano	gita	raʔit
3P	pano	gita	raʔit

With a few heavily used verbs, the result has been outright irregularity (ibid., 607):

TABLE 8.11. “Eat” in Sika

	<i>gea</i> “eat”
1S	ʔoa
2S	gea
3S	ga
1P (exclusive)	gea
1P (inclusive)	ʔea
2P	gea
3P	ʔa

In Dawanese (aka Timorese) of Tetun’s Fabronic group, there are two sets of subject prefixes, syllabic ones and ones consisting of only a single consonant, with it unpredictable which set a given verb takes (Steinhauer 1993, 133–34). Verbs change shape in several fashions (including metathesis) according to factors such as presence and placement of object and whether a following object begins with a consonant cluster (ibid., 141–42). Thus with no object:

- (55) Qau qòt  
I cut  
‘I cut.’

With a fronted object:

- (56) Nuif qi qau qote  
bone this I cut  
‘This bone I cut.’ (142)

The same shape occurs if the following object begins with a cluster:

- (57) Hò m-qote qfulaq.  
you 2S-cut flower  
‘You cut flowers/a flower.’

But otherwise, a following object triggers metathesis:

- (58) Hò m-qoet nuif?  
you 2S-cut bone  
‘Do you cut bones?’ (141)

Then there are irregular verbs (ibid., 154–55):

Table 8.12. Irregular Verbs in Dawanese

	subj. pron.	“come”	“eat”
1S	qau	qóm	qua
2S	hò	qóm	mua
3S	qin	nèm	nah
1P (exclusive)	hai	qém	mia
1P (inclusive)	hit	tèm	tah
2P	hi	qém	mia
3P	sin	nèm	nah’an

Thus TD stands out among its relatives in terms of overspecification, complexity, and irregularity.

### 12.2. *Similar Cases: Timor and Flores Island*

However, TD is not the only language variety in Central Malayo-Polynesian that is saliently simplified morphologically. On Timor, Tokodede of the Ramelaic sister group to Fabronic is as reduced morphologically as TD. Meanwhile, there are a few languages on the island of Flores, westward of Timor, that have no morphology whatsoever, derivational or inflectional. The ones currently known to me are of the Bima-Sumba group: Kéo (Baird 2002) and Ngadha (Djawanai and Grimes 1995).

If these languages had grammars like this simply by chance, then they would be counterexamples to a general claim of this book, that languages do not shed all or most of their inflections as a normal, internal development, and that this can always be traced to heavy non-native acquisition at some point in the past. But in fact, the languages in question share a unique sociohistorical factor.

There is a unique and strong lexical and grammatical imprint in languages of Timor and Flores from Central Malayo-Polynesian languages spoken north of Timor, on the Uliasser islands—Ambon and adjacent islands—and to a lesser extent, on nearby Ceram, Buru, and Ambelau. The languages include Asiluluan, Hitunese, Larikean, Batumeran, Harukuan, Saparuan, and Bahasa Nusa Laut of the Uliassers, Nialan, Bonfian, Warunese of Ceram, and Masarete and Kayeli of Buru (Hull 1998, 154). Hull refers to this source generically as an “Ambonic” imprint.

The lexical legacy in Timor includes about a hundred items, of core semantics rather than “cultural” borrowings, including grammatical items such as “no,” prepositions such as “at” and “to,” and adverbs such as “before” and “tomorrow” (Hull 2001). Languages of Timor also have various grammatical features found in the Ambonic languages but not in the Sulawesi languages from which the Timoric languages developed (of the Muna-Buton group that Tukang Besi and Muna belong to), such as an inalienability distinction, an agentive prefix, a negative imperative marker, postposition of numerals and preposing of possessive markers (Hull 1998, 166; Hull 2001, 115).

This impact, extending to core lexicon and grammar, is indicative of extensive non-native acquisition of earlier stages of these languages, rather than mere passing contact resulting in marginal borrowings. The extent and depth of the lexical borrowings, for example, are analogous to the Scandinavian ones in English. Hull (1998) reconstructs that in approximately the 1200s, invaders from the Ambonic-speaking region disrupted language transmission where they settled. Evidence beyond the linguistic includes that six place names in Ambon correspond to the names of tribes of Timor; that Timorese origin myths correspond to lore connected with the Moluccan aristocracy (Ambon and surrounding islands are part of the Molucca group) rather than to that connected with Sulawesi nobility; that Rotinese tales of origin describe immigration by a band of Ceramese over four hundred years before the arrival of the Portuguese; and that speakers of the Timoric language

Galoli recount that in antiquity, men from Nusa Laut arrived and married local women (ibid., 161–64).

The Ambonese imprint accounts for the otherwise bizarre grammatical profile of the languages in question. Besides the precise correlation between this imprint and the grammatical simplification, two other factors point clearly to a contact explanation.

#### 12.2.1. TIMOR AND FLORES BEYOND TETUN DILI AND THE BIMA-SUMBA CASES

First, TD can be seen as an extreme of a general tendency toward, especially, unusual morphological simplification in Austronesian languages of Timor and Flores. The highly conventionalized focus- and/or voice-marking apparatus diagnostic of Austronesian is absent, and not transformed into or replaced by any similarly elaborated feature (from Hull 2001, 163):

Table 8.13. *They stole my buffalo and killed it* in Four Timorese Languages

Tetun Terik	Sira	na'ok	ha'u	karau,		roho	tiha.
Galoli	Sia	na'o	ga'u-ni	karau	no	regen	oin.
Baikenu	Sin	nabák	au	bijae,		nlol	nalali.
Mambai (Ainaro)	Rom	pnao	au	arbau,		dac-pul	tel.
	they	steal	my	buffalo	and	kill	PERF

Philippines-type languages would require focus marking via both affixation and free morphemes on a constituent; in these languages, focus is marked only for emphasis. Most IT languages have concordial marking for subjects and/or objects; among these four languages, only Baikenu has a third-person plural subject affix. Even Malay has grammaticalized voice marking, but this is absent in all of these languages.

This underspecified tendency occurs according to degree. Contrasting with relatively conservative languages like Rotinese and Kambara are intermediate cases such as Manggarai, which has no subject-marking prefixes but has paradigms of object-marking and possessive clitics (Verheijen and Grimes 1995).

But because conservative languages like Rotinese and Kambara exist, we can assume that this drastic morphological simplification does not trace to an ancestral language to those of Timor and Flores. Rather, it would appear that the erosion has happened after the emergence of the languages, to varying degrees. Claiming that this was a mere matter of chance requires pulling the camera back and thinking of Austronesian in its full spread: where else in the group has chance had this effect?

Importantly, the Ambonese encounter also answers a possible question. Tetun Dili is, after all, not the only Austronesian language besides Malay used as a lingua franca. Javanese and even Tukang Besi are examples, and yet they are hardly as simplified as TD. Like the PMD Malays, TD is based on an already relatively simplified



precursor: TT itself, and its close relatives, are strangely simplified compared to the typical Austronesian language.

#### 12.2.2. THE ANOMALY OF THE PAPUAN-RELATED LANGUAGES OF TIMOR

Second, the four Papuan-related languages of Timor — Bunak, Makasai, Makalero, and Fataluku — show the same simplification, despite their Papuan relatives' retaining typically synthetic grammars. Makasai, in particular, is as morphologically stripped as TD. This clearly refutes a mere genetic explanation, in that these languages do not descend from earlier stages of Central Malayo-Polynesian.

Of course, we might attempt to preserve the chance account in the Timorese and Flores languages by supposing that the Papuan-related languages developed in this direction because of contact with the reduced Timorese languages. But this would ascribe a highly unusual degree of structural loss to the contact in question. To be sure, Sprachbund effects can occasion loss modeled on one or more of the languages in contact, such as the cases noted in section 11.2.2, or the loss of concordial morphemes that I propose were modeled on Malay in Malayo-Javanic. But in all of these cases, a great deal of elaboration was retained: the Balkan languages remain highly inflected, as does Gurnu Baagandji, and the languages of Java retain their derivational morphology and certain inflections.

But in the case of the Papuan languages of Timor, we must explain why these four languages have not just lost some inflections of a particular kind, but why they lost so much morphology overall that they stand as strangely analytic, or analytic-leaning, languages in sharp contrast with hundreds of synthetic languages of Papua New Guinea. A mere contact account becomes especially incoherent in view of the fact that so very many languages of Papua New Guinea and islands eastward have been spoken alongside isolating languages like Tok Pisin and its sister creoles for a century and a half.

If we suppose that the effect would require more time, then other contrasts would include the Bantu languages of South Africa. Contact starting as long as three thousand years ago with Khoi-San languages was intimate enough to occasion the borrowing of features as marked as click phonemes, and yet Bantu languages of the region retain the group's heavy agglutination of affixes on verb roots leaning toward polysynthetic typology, including object affixes, rather than showing any sign of Khoi-San's variably inflected but much less concatenative morphology. Agglutinative Altaic languages have been spoken alongside Mandarin for millennia, and yet the result is not analytic Altaic languages but agglutinative hybrids of Mandarin and Altaic.

That being the case in Altaic languages and Mandarin, why, then, are there not at least one or two of the four Papuan-related languages of Timor in which Papuan morphology remains, but is remodeled partially upon the morphological categories of conservative varieties like Tetun Terik? Why not Papuan languages where most of the morphology remains, but only slightly eroded by contact?

Rather than assume that mere serendipity is responsible here, it is more grace-

ful to assume that a larger factor is responsible for the unusual simplification in all of the languages of Timor and the ones in Flores, especially given that a single sociohistorical phenomenon correlates with precisely the area in question. This is especially attractive given that cross-linguistically, it is not “normal” for languages to “undress” for no reason. Short of the discovery of mysteriously analytic languages in Algonquian, Australia, or the Caucasus, a contact explanation of the Timor-Flores case is required by Occam’s razor.<sup>8</sup>

### 12.3. *Timor and Flores and the Languages of the World*

The anomaly of Tetun Dili and languages like Tokodede and Kéo is even clearer when we realize that these grammars are peculiar in the cross-linguistic sense.

The vast majority of the world’s languages have either inflectional morphology or tones that are either lexically or morphosyntactically contrastive. Only a few languages that are not creoles lack both: to my knowledge, these are the Polynesian languages, some Mon-Khmer ones, and, crucially, Florinese languages like Kéo and Ngadha. But even in these languages, there are other features that either fulfill functions usually carried out by inflections or are phonetically and functionally analogous to tone or related to its life cycle.

Polynesian languages have no inflections or tones, but have free morphemes that indicate grammatical roles, including ergativity:

- (59) ‘oku tamate ‘e he fefine ‘a e tangatá.  
 PRES kill    ERG DEF woman ABS DEF man  
 ‘The woman kills the man.’ (Tchekhoff 1979, 409)

There are Mon-Khmer languages with neither inflection nor tone, but most of the ones without tone have register distinctions (i.e., between clear, breathy and/or creaky voice) whose production is highly similar to that of pitch distinctions and often difficult to distinguish from it, and carry out a similar grammatical function, as in Bruu of Thailand (Thongkum 1981, 230–31):

- (60) Minimal pairs distinguished by register in Bruu  
 tuup “to be thick (e.g., of cloth)”  
 tùp “to bury”  
 luup “to leap over”  
 lùp “to pat fondly”

There are Mon-Khmer languages with neither register or tone. But the development of registers also conditions a proliferation in vowel distinctions, which remain after register falls away. Mon-Khmer, in fact, has the largest known vowel inventories in the world, many languages having as many as three dozen distinctions (including phonemic length-based ones). Thus when register disappears in Mon-Khmer languages, it does not leave behind a mere seven to ten vowels, but a calling card in the form of uniquely vast vowel arrays. Khmer has, for example, no register or tone but thirty vowels (Ferland 1992):

TABLE 8.14. Khmer Vocalic Inventory

iə	ɨə	uə			
ii	ɪɪ	uu			
ee	əə	oo	e	ə	o
ɛe	ʌə	ɔo			
ɛɛ		ɔɔ	ɛ	ʌ	ɔ
aɛ	aʌ	aɔ	ɛ̃a		ɔ̃a
	aa	aa	a		

Meanwhile, while Polynesian languages have their particles corresponding to functions often carried out by inflection, Mon-Khmer languages also have what can be viewed as functional equivalents to gender inflection. As members of the South-east Asian Sprachbund, these Mon-Khmer languages have ample numeral classifiers. Khmer, for example, has various ones corresponding to animacy, shape, and so forth, such as *səwphɨw muəy lou* “a dozen books,” *kmaw-day pram daəm* “five pencils,” *puəq-maaq pii nəəq* “two friends” (Huffman 1970, 69–70).

Nor is grammatical gender itself alien to Mon-Khmer. Khasi of India and Bangladesh has only two classifiers, *ɣut* for humans and *tilli* for non-humans (Nagaraja 1985, 79). However, it has masculine and feminine grammatical gender particles that pattern broadly according to type and shape of object and assorted classes of abstract noun, but with a substantial degree of arbitrariness as well, paralleling the situation with gender marking in European languages (Rabel-Heyman 1977). Some examples:

TABLE 8.15. Grammatical Gender Markers in Khasi

masculine		feminine	
<i>ʔu tmem</i>	“hammer”	<i>ka makhan</i>	“butter”
<i>ʔu prthat</i>	“thunder”	<i>ka ʔiktɨar</i>	“authority”
<i>ʔu khllaay</i>	“kidney”	<i>ka khroŋ</i>	“tax”
<i>ʔu kubi</i>	“cabbage”	<i>ka knta</i>	“hour”
<i>ʔu ʔata</i>	“flour”	<i>ka kurat</i>	“saw”
<i>ʔu braw</i>	“vessel”	<i>ka snɨer</i>	“intestine”

Here, then, is the puzzle regarding Tetun Dili, similar Timorese grammars, and the Flores languages. Tetun Dili has a remnant of inflection if we consider its intransitivizing prefix an inflection, but languages like Kéo and Ngadha are completely analytic. Thus only their use of numeral classifiers parallels the developments in Polynesian and Mon-Khmer, as in Kéo, in which Baird (2002, 252–53) lists twenty-four sortal classifiers (as well as several others used exclusively with one nominal):

- (61) Wondo ndua,    mbu’ɛ    **ngga’e** dima    rua    mbana kima  
Wondo go-down    girl    CL    five    two go    mollusc  
‘Seven young Wondo women went down to collect molluscs’ (Baird 2002, 490)

These languages have no free morphemes regularly tracking the grammatical function of dependents, no register distinctions, and nothing even resembling the vowel proliferation of Mon-Khmer. It is extremely unusual for an older language to hover this closely above creoles in terms of overspecification and irregularity. That is, I am aware of no such cases other than, precisely, Tetun Dili, the few other relevant languages of Timor and Flores, and, pointedly, Riau Indonesian and similarly reduced Malay varieties—anywhere in the world, among the six thousand languages extant.

The contrast between Tetun Terik and Tetun Dili is unequivocally due to non-native acquisition. We might ascribe the other cases to an exotic happenstance in just these languages out of six thousand. But a more scientific, systematically oriented perspective would seem to all but require attributing the anomaly of Kéo and Ngadha to contact as well. That is, this grammatical typology is so glaringly unusual cross-linguistically beyond creoles—established as products of non-native acquisition—that it is most graceful to assume that its appearance in a language without a written history must be due to contact as well.

### 13. Conclusion

To assume that some Austronesian languages have become drastically less complexified simply by chance, for the same random reason that French eliminated the simple past in favor of the perfect construction, makes a simple prediction. That prediction is that scattered randomly among the one thousand Austronesian languages, there will be at least dozens of fortuitously analytic and underspecified SVO grammars, contrasting with ones spoken next door retaining inflectional paradigms, four-way demonstrative gradations, and VSO word order.

This is not remotely the case. There are but a handful of outlier grammars of this kind, and in all cases, contact accounts distinguish their histories. Malay has been one of the world's most widely spoken *lingua francas* for most of its existence. The possibility that what Dalby (1998, 391) terms its “more approachable structure than its relatives” is unrelated to this, or not closely related enough to merit extended address, is belied by the fact that another oddly simplified Austronesian language, Tetun Dili, is another longtime *lingua franca* sitting alongside a considerably more conservative variety rarely acquired by outsiders. The case is bolstered even further by the fact that the handful of nearby languages paralleling Tetun Dili in simplification share a decisive and unique imprint from languages brought to the island by speakers of another one.

A systematic account of a wide range of data must allow breathing room for the operations of chance. But in the case of Malay, the starkness of its typological contrast with its relatives combines with unique aspects of its sociohistory to force a causal account. Malay is another example of how long-term non-native acquisition has decisive effects upon the transmission of older languages.

# A New Typology of Language Contact

I hope that I have convinced the reader that there is a qualitative difference in overall structural complexity between five languages and their sisters, and that this difference was caused by incomplete acquisition of a moderate degree in contact situations.

If I have been successful, then a problem remains: that there is no clear place for the phenomenon I have identified in conventional models of language contact. Instead, the reduction in question, where perceived at all, has tended to elicit surmises that the languages are, or were once, pidgins or creoles. But these claims fail to compel, because neither synchrony, diachrony, nor sociohistorical data support classing these languages with Saramaccan, Haitian Creole, Papiamentu, and Tok Pisin.

## 1. The Pidgin/Creole Detour

The problem is that pidgins and creoles are typically taxonomized in ways that discourage awareness that pidginization and creolization are clinal phenomena. First, pidgins and creoles are typically classified separately from other cases of language contact, partly because of a certain taxonomic unity suggested by most of them having arisen amid plantation slavery during four centuries under a few Western European powers.

Thomason and Kaufman (1988), for example, treat creoles as peculiar examples of incomplete transmission, separately from their other chapters on borrowing and interference. While they submit the former processes to a careful scalar schema, they only address in passing that incompleteness of transmission is also clinal, such as a brief reference (155) to the difference between Mauritian Creole French and Réunionnais's "semicreole" French. Where they discuss Afrikaans in a brief case study, they emphasize that it is "not a creole," but only to place it as a case of language shift similar to the many other cases they discuss in the book. While they describe in it a degree of structural reduction larger than that in most of their other

case studies, their classification of Afrikaans with the other language shift cases in the book detracts attention from this unusual degree of reduction as a key distinction. In other works (e.g., Thomason 1997, 84–86), it is clear that Thomason is aware of the semicreole class and the inherently fuzzy boundaries between types of language contact. But in the 1988 book, the precise nature of the zone between what the authors term “abrupt creolization” and what they term “language maintenance” is unclear, partly because of the authors’ primary interest in the language maintenance cases. The general impression is that creoles, as products of a break in transmission, are rather like a disconnected section of a worm’s body, which stays alive and becomes a new worm that wriggles off by itself.

Then second, since the 1990s a fashion in creole studies has proposed that creole genesis is simply a matter of second-language acquisition amid heavy mixture, unconnected from the unequivocal break in transmission that is pidginization. This framework, resisting the idea that creolization involves any significant break in transmission, inherently leaves no room for qualitatively differentiating the emergence of Saramaccan from Modern English at all. Rather, authors such as DeGraff (2001) imply that, properly speaking, it is scientifically inappropriate to even consider creoles to constitute a class of language.

Taxonomizing the languages in the case studies in this book will require getting past both of these epistemological barriers.

First, pidgins and creoles are indeed a phenomenon worthy of address as a class, but only in being extreme ends of a continuum of structural reduction, upon which various languages could be placed at each point in an unbroken line.

Meanwhile, I suspect that the conception of creoles as merely mixed rather than simplified is largely embraced by certain subschools within the creolist community. Most linguists readily process a qualitative difference between Romanian and Saramaccan other than that they were born from different source languages. The impression of extensive grammatical simplification that creoles give upon initial observation is not illusory, as I have argued elsewhere (cf. especially the first five chapters of McWhorter 2005), and arguments otherwise are founded upon sociopolitics rather than empiricism.

Thus there is an inherent incompatibility between the foundations of my taxonomy and those of the school seeking to abolish creole studies as a discipline. My subsequent discussion will proceed with an assumption that this school of thought has been refuted; those who remain convinced of its tenets will disagree totally with the rest of this chapter.

## 2. Bridging the Gap

My interest is in fashioning a discrete place within language contact taxonomy for languages that have been slightly decomplexified by non-native acquisition, but are not and never have been pidgins or creoles.

The basic claim is that in language contact, mixture and reduction operate on complementary axes. After contact, some languages become sharply reduced with little or no mixture beyond the inevitable lexical borrowings. In contrast, some lan-

guages become highly hybridized not only in lexicon but also in all levels of grammar, but without being structurally reduced to any significant degree. Meanwhile, in most cases, the result of language contact exhibits both reduction and mixture to varying degrees.

If this claim is true, then we expect that there will be examples of cases exhibiting all possible degrees of this mixture, including those where either reduction or structural mixture is absent. I believe that this is true, and illustrate it in the following table. The horizontal axis represents degrees of mixture, from the solely lexical to the deeply structural. The vertical axis represents degrees of reduction, from essentially none to the vast.

Table 9.1. A Schema of Language Contact Results

	lexicon only	lexicon and syntax	lexicon, syntax, morphology, phonology
no simplification	1. German in Finnish	2. Romanian and other Balkan Sprachbund languages	3. Media Lengua Ancient language areas (Amazon, Australia)
some simplification on all levels	4. English, Persian, Malay, Mandarin, New Arabic, AAVE, Popular Brazilian Portuguese	5. Afrikaans, Réunionnais French, Singapore English	6.  Shaba Swahili
extreme simplification on all levels	7. Hawaiian Creole, Chinook Jargon creole	8. most creoles	9. Saramaccan, Tok Pisin, Berbice Creole, Dutch

One might suppose that this schema is a mere tabulation of assorted outcomes, representing no systematicity other than correspondences between particulars of contact situations (demography, social relations, language types, etc.) and the new varieties that result. But in fact, each row encompasses languages commonly classed together, or that have occasioned similar controversies and confusions.

3. Row 1

The first row can be thought of as the “Thomason and Kaufman” row, representing their cline of intensity of borrowing. This row also encompasses what they term the alternate case of interference, when non-natives’ acquisition of a language retains varying degrees of influence from their native languages. I consider the cen-

tral insight to be what happens to a particular language under impact from another one, regardless of whether the influence occurs when native speakers incorporate foreign elements into their grammar or foreigners develop a version of the grammar incorporating features from their own. Thomason and Kaufman's distinction between borrowing and interference is intended to guide linguists in reconstructing the contact situation a language emerged in; for example, Romanian *borrowed* from Balkan languages, but Amharic resulted when Cushitic speakers left *interference* in their acquisition of Semitic. However, this distinction is not necessary to a schema addressing degree of contact influence in a language rather than how that contact occurred.

### 3.1. Box 1: Lexical Borrowing

Box 1 is mere lexical borrowing, including even extensive cases. Finnish, for example, gives no indication of especial simplification, but has a large lexical inheritance from German.

### 3.2. Box 2: Moderate Sprachbund Influence

Box 2 represents cases where borrowing, under Thomason and Kaufman's definition, extends to syntax but not the borrowing of phonological features or morphology (on their scale, basically levels 2 and 3). Romanian's transformation within the Balkan Sprachbund is typical. Most of the impact here was syntactic, such as the postposition of definite articles (*om-ul* "the man") and use of pleonastic pronouns cross-indexing direct and indirect animate objects:

- (1) I-am scris lui Ion.  
to-him I.wrote him John  
'I wrote to John.'

Elsewhere were elisions of morphology in favor of analytic constructions in other languages in the Sprachbund, such as the development of a periphrastic future, the loss of infinitival constructions in favor of finite ones, and the merger of dative with genitive case. These affected Romanian morphology, but did not entail the borrowing of morphological structures or morphemes from the other languages. Meanwhile, most of the phonological influence is subtractive, such as loss of length distinction; additions, such as of phonemic central vowels, are minimal.

Moreover, to the extent that these instances led to a certain amount of simplification, it was a mere drop in the bucket. Elsewhere, Romanian exhibits various features more complex than most other Romance languages', such that it is in no sense a less complex language than them, and is arguably more complex. Examples include the redevelopment of a neuter gender (that patterns with masculine in the singular and feminine in the plural); the retention of a genitive/dative case marker (this complicated by its expression only with the determiner in the masculine and neuter); frequently wide phonetic distance between free and cliticized pronominals in the dative and accusative (dative first-person singular free *îmi*, clitic *mie*; accusative



third-person singular *îl* free, *o* clitic); especially complex plural marking according to gender and phonetic classes within them; and so on.

### 3.3. Box 3: *Intertwined Languages and Strong Areal Influence*

Box 3 contains extreme cases of borrowing that extend to morphology and phonology. This includes the languages called “mixed languages” or “intertwined languages,” such as Media Lengua, Mednyj Aleut, and Ma’a. Media Lengua, for example, inserts Spanish morphemes into Quechua morphosyntax and phonology; note in the example below that the Spanish lexical items are filtered through Quechua’s lack of mid vowels:

- (2) (a) Media Lengua  
 Unu fabur-**ta** pidi-**nga-bu** bini-**xu-ni**.  
 a favor-ACC ask-NOM-BEN come-PROG-I  
 ‘I come to ask a favor.’
- (b) Spanish  
 Vengo para pedir un favor.  
 come.I for ask a favor
- (c) Quechua:  
 Shuk fabur-da maña-nga-bu shamu-xu-ni.  
 one favor-ACC ask-NOM-BEN come-PROG-I (Bakker and Muysken 1995, 43)

Box 3 also includes language areas with rampant sharing of not only syntactic but also phonological and inflectional structure, the latter via either remodeling native morphemes on the behavior of those in Language B or borrowing the morphemes themselves. This kind of mixture results from what Dixon (1997) terms *linguistic equilibrium*, in which languages “stew” together over long periods of time due to contact of varying degrees, becoming ever more alike typologically.

Here fit cases such as the Australian mixture described by Heath (1997), or the transformation of the Arawak language Tariana in the Vaupés basin language area under influence from East Tucano languages. Tariana underwent syntactic changes, such as to verb-final word order and frequent dependent marking rather than Arawak’s head marking, phonological changes, such as nasal prosody and pitch accent patterns verging on tonal distinctions, and extensive inflectional remodeling in case markers, classifiers, evidentials, tense, switch-reference marking, and other categories, such that the totality of the changes render Tariana much closer typologically to East Tucano than to Arawak (Aikhenvald 1999).

## 4. Row 2

The second row represents the nut of this book. These are cases where language contact has resulted in moderate simplification, not extreme enough to classify as pidginization or creolization, but curious enough to have motivated various propos-

als that something irregular occurred in their development. Languages in row 2 have commonly occasioned controversies as to whether they were creoles or examples of normal internal development, when, in fact, they are something intermediate.

#### 4.1. Box 4: Our Friends from the Case Studies

Box 4 contains the five languages treated in the case studies. Contact has rendered them somewhat less complex than their sisters—but without structural mixture being a significant factor. Rather, mixture has been overwhelmingly lexical. Thus it is not surprising that there is no significant French grammatical influence in English nor non-native grammatical influence in the New Arabics; Altaic structure plays a very small part in distinguishing Mandarin from other Chinesees; and Malay differs from its relatives for reasons beyond its membership in the Sinosphere Sprachbund (given that all languages of the region are members of that Sprachbund as well). Structural reduction can occur without grammatical mixture, and this book has attempted to demonstrate this in the five case study languages. Dahl (2004) is a rare example of a language contact researcher who readily acknowledges this. In response to Thomason and Kaufman's (1988) general implication that language shift is a largely a process of transfer from L<sub>1</sub> into L<sub>2</sub>, Dahl notes (281):

Imperfect learning need not involve any direct influence of the other language. Commonly, what happens is simply that the learner fails to pick up some feature of the target language—this is the filtering effect of second-language learning. . . . Instead of interference in language shift we should therefore rather speak of sub-optimal transmission.

#### 4.2. Box 5: Semicreoles

Box 5, meanwhile, contains the languages often termed *semicreoles* in the creolist literature. Creolists such as Holm (2003) have observed that these languages exhibit both structural reduction and mixture to a degree intermediate between full transmission and creolization.

##### 4.2.1. SOME EXAMPLES

Afrikaans, as argued in chapter 4, has structurally abbreviated the Proto-Germanic legacy no more than English: thus its membership in the same row as English. However, it does have some signs of structural mixture from Khoi and varieties of Malay spoken by subordinate peoples amid the language's birth. These include (data from Holm 1989, 346–49) double negative marking (*Hy gaan nie hieso kom nie* “He will not come here”), possession indicated with a genitive particle postposed to pronouns (*julle se vriende* “your friends”), and reduplication of nouns and verbs (*Die mense daar het troppe-troppe beeste* “People here have herds and herds of cattle”) in contrast to the restriction of the feature to adjectives and adverbs in Dutch.

Réunionnais French similarly reduces standard French structure to a degree too moderate to treat as full pidginization or creolization, but much more than even the

most nonstandard varieties of French. Examples include virtual eclipse of grammatical gender, elimination of the inflectional plural and most conjugational allomorphs, and reduction of case distinction in pronominals. Meanwhile, it bears some signs of the grammars of its Malagasy and Bantu creators, such as in aspects of coordinate, consecutive, and serialization constructions (Corne 1999, 85–96, 181–96) and in that fossilized agglutination of French definite articles in nouns (*lapo* “skin” < *la peau*) is more common in Réunionnais (and nearby Mauritian French Creole also created by Bantus) than in French creoles elsewhere, likely due to congruence with Bantu noun class prefixes (Baker 1984).

A final example here is Colloquial Singapore English or “Singlish.” The language is heavily influenced by Mandarin Chinese, but mostly in terms of syntax. A sentence like *Is very interesting I find geography* reflects Mandarin’s topic-comment typology, *We don’t have* its tendency to leave arguments unexpressed, *Take come here, can or not?* as a command to bring a book Mandarin’s serial verbs (Ho and Platt 1993, 18). But there is no reflection of Mandarin features expressed morphologically in many languages, such as the delimitative aspect construction expressed with reduplication and *yī*:

- (3) Wǒ xiǎng kàn yī kàn nèi-jīān wūzi.  
 I want see one see that-CL house  
 ‘I’d like to have a look at that house.’ (ibid., 181)

(\**He wants to see one see that house.*) There are no equivalents to numeral classifiers. Phonological differences from standard English are mostly subtractive symptoms of incomplete acquisition such as absent final consonants, contraction of consonant clusters, and syllable elisions (Platt 1975, 370); in no sense is Singlish phonology significantly parallel to particulars of Mandarin’s. (See Bao 2005 for a similar argument regarding Singlish’s chariness of incorporating Mandarin features that are especially divergent from English.)

Meanwhile, Singlish belongs in row 2 because its reduction is robust but not to the point of pidginization. English inflections tend to be expressed variably; for example, past marking is retained 57.3 percent of the time when expressed with ablaut, but only 3.9 percent of the time when final *d* is part of a consonant cluster (*told*), and at rates in between for other allomorphs of the *-ed* inflection (ibid., 75). (Nor is this a matter of variable competence along the continuum from the basilect to standard English; this ordering frequency holds across educational levels [ibid., 93].) There is also variable omission of the verb “to be.” While Ho and Platt attribute this to the fact that Mandarin’s equative copula *shì* is variably eliminated, this does not explain why Singlish also often omits the copula in locative contexts, despite the fact that Mandarin’s locative copula *zài* is obligatory. The guiding factor would seem to have been simplification, given that copular omission is a virtual diagnostic of non-native acquisition of grammars.

#### 4.2.2. TAXONOMIC HOUSECLEANING

Readers familiar with the literature on semicreolization will note that I class African American Vernacular English (henceforth AAVE) and Popular Brazilian Portuguese,

often classed as semicreoles, in box 4 rather than box 5. This is because these languages exhibit virtually no structural features from other languages.

Popular Brazilian Portuguese contrasts with European Portuguese and Standard Brazilian Portuguese most saliently in reduction of morphological allomorphy. In the present, *-o* occurs in the first-person singular and *-e* in all other cases. Concordial agreement is variably but vastly reduced (*as casas brancas* > *as casa branca*). In especially reduced dialects such as in Helvecia, there is no allomorphic variation at all in preterite marking (De Mello et al. 1998). These and a few other traits place Popular Brazilian Portuguese in our second row.

There is a tendency among scholars of this language variety to trace features like this to the analyticity of West African Kwa languages like Yoruba. But this neglects that Palenquero Creole Spanish, created almost exclusively by speakers of highly agglutinative Kikongo, is as analytic as most creoles. This is one more indication that structural reduction is often a central factor in language contact regardless of the features available to mix in.

Crucially, otherwise Popular Brazilian Portuguese has few structural aspects traceable to other grammars at all. Holm (1987, 416–17) notes residual evidence of an imposition of CV syllabic structure upon Portuguese words, such as variable occurrences of *nego* for *negro* “black person” and archaic attestations such as *lapassi* for *rapaz* “boy.” These could be modeled on Kwa languages’ similar phonotactics, but would also be predictable as results of non-native acquisition, since, for example, there is a similar tendency toward CV phonotactics in the Shaba Swahili lingua franca as well as Kenyan Pidgin Swahili, despite the fact that standard Swahili and the related Bantu languages spoken by these varieties’ creators were all already strongly CV in their phonotactics (De Rooij 1997, 319–20). Then, the use of *ter* “to have” as a verb “to be” (*Lá no céu tem uma estrela* “There in the sky is a star” [Holm 1987, 419]) could well be modeled on Indo-Portuguese creoles’ similar feature, but then 1) this could be treated as a lexical rather than grammatical inheritance, and 2) it is also a cross-linguistic tendency in existential sentences to use *have* verbs, such as even colloquial English *They’ve got eight new stores down there since last year*. More convincingly traced to outside sources is the double negation, as in *Não quero não* “I don’t want to,” paralleled by double negation in Portuguese creoles of Africa like that of São Tomé (*Na kâtá-fa* “Don’t sing”) (ibid., 414). But I am unaware of a single other structural feature in Popular Brazilian Portuguese that parallels foreign ones to this conclusive degree.

Meanwhile, African American Vernacular English is similarly devoid of features traceable to other languages, as I have argued in McWhorter (1998, 2000c, 419–24). African American Vernacular English differs from standard English largely in reductions. These include zero-copula, as in *She my sister*. Claims that this is an African feature are based on the surface parallel between predicate adjective sentences such as *He big* and Fongbe *Sáki ó wí* [bag DEF black] “The bag is black” (Lefebvre and Brousseau 2001, 358). But this neglects a larger picture. West African languages usually encode what are adjectival concepts in European languages with verbs, as is common worldwide, such that, properly speaking, they exhibit no absent copula in such sentences, which translate technically as “He bigs.” And then elsewhere in their grammars, West African languages offer no model for zero-copula in equative and locative constructions, in which it is in fact a typological diagnostic to have separate

copulas for both contexts: Ûn *nyí* Àfíáví ‘I am Afiavi’; Wémâ s̄ d̄s̄ távò jí [book DEF COP table on] ‘The book is on the table’ (ibid., 144, 147).

Meanwhile, it is difficult to treat elimination of subject-auxiliary inversion (*Why you don’t call me anymore?*) as an “African” feature when this kind of inversion is extremely rare outside of Europe (Ullan 1978), and thus is plausible as a simplification predictable of non-native acquisition. The moderate degree of inflectional reduction such as the bare third-person singular form (*he walk*) and possessive (*Billy book*) has models in nonstandard British varieties, and, overall, AAVE remains as inflectionally rich as standard, including strong verb ablaut (*took, thought*) and irregular plural nouns (*men, people, children, geese*). Possibly, the elimination of syllable-final *r* (*sto’* for *store*) is traceable to West African CV phonotactics, since this feature had yet to become regular in nonstandard Englishes in the eighteenth century when AAVE likely developed. But this alone cannot motivate classing AAVE with more prevalent cases of mixture such as Afrikaans and Réunionais.

#### 4.3. Box 6: Semicreoles Based on Typologically Close Languages

Box 6 represents cases in which reduction has been slight, but mixture has been more intimate than in the semicreoles. Shaba Swahili, a vernacular variety developed by speakers of various Bantu languages in mines in the Katanga province of the Congo between 1920 and 1940, is an example. Typically, it has been proposed that Shaba Swahili was a “creole” (Polomé 1968), followed by an objection (Kapanga 1993) that there is no evidence of a pidgin Swahili in Katanga and that it is not as reduced compared to Swahili as creoles are to their lexifiers. Kapanga concludes that Shaba Swahili simply eliminates features most divergent between Swahili and the languages its creators spoke, implying that it is a koineization along the lines of that of Attic Greek, but not a simplified variety of Swahili in any significant sense.

But De Rooij (1997) demonstrates that Shaba Swahili evidences both mixture and structural reduction, hardly as extreme as that in pidgin varieties like Kenyan Pidgin Swahili, but enough to distinguish it from what we would expect of Swahili transmitted without interruption.

Thus contributions from the languages of the miners who created it, such as Bemba, Luba/Kasai, Songye, Kanyok, and Kete, are clear. Where standard Swahili encodes relativization with an affix, Shaba Swahili has a postposed free morpheme agreeing with the referent, as is common in the substrate languages (SP = subject prefix; CP = concordial prefix):

- (4) (a) Standard  
 mtu a-si-ye-soma  
 man NEG-PAST-REL-read  
 ‘a man who doesn’t read’
- (b) Shaba  
 Ni-ta-ku-pa mutoto u-le ni-ta-zala  
 I-FUT-you-give child CP-REL I-FUT-bear  
 ‘I will give you the child I will give birth to.’ (De Rooij 1997, 328–29)

The mixture also extends to the inflectional module, as Shaba Swahili actually has three more noun classes than standard, having inherited these from the substrate languages (Kapanga 1993, 447).

Kapanga treats the latter feature as evidence that there is no complexity differential between Shaba Swahili and the standard. But De Rooij shows that this neglects a general reductive tendency in Shaba Swahili, such that the three new noun classes no more contradict a reductive analysis than does *do*-support in English.

Shaba Swahili lacks the final nasal + consonant clusters in some of the substrate languages such as Kete, in favor of maintaining as much as possible a CV phontactics (De Rooij 1997, 319). In the standard, adjectival concord varies according to noun class, but in Shaba Swahili adjectives usually have one (or two) generalized forms (CL = noun class prefix):

- (5) (a) Standard: **m**-toto **m**-dogo a-li-soma **ki**-tabu **ki**-zuri.  
 (b) Shaba: **mu**-toto **ki**-loko a-ri/li-soma ki-tabu **ki**/**mu**-zuri  
 CL-child CP-small SP-PAST-read CL-book CL-good

‘The small child read a good book.’ (ibid., 322)

Shaba Swahili overall tends to reinterpret the idiosyncracies of the standard’s noun class taxonomy according to a simpler distinction between animate and nonanimate. Animates that in the standard occur in nonanimate classes are brought into animate classes 1 and 2: for *bwana* ‘gentleman,’ where in the standard it occurs in the *ji/-ma*-classes 5/6 (*bwana/mabwana*), in Shaba it is placed in the more ‘logical’ animate 1/2 classes (*bwana/babwana*) (ibid., 321). The subject marker is usually invariantly *i*- with nonanimates, rather than varying according to the various noun classes as in the standard (ibid., 322–23). Standard Swahili cross-indexes full NP arguments with object affixes on the verb, but in Shaba Swahili this is rare with inanimates in all but one class (ibid.). In general, as De Rooij (330) appropriately concludes, in Shaba Swahili simplification ‘has gone further than might be expected if selective simplification had been the only restructuring process.’ To the extent that no Bantu language without extensive rapid non-native acquisition in its history is presented that parallels Shaba Swahili in this degree of reduction, we must accept that Shaba Swahili’s contrasts with the standard represent a degree of simplification unattributable to uninterrupted internal change.

Other languages in box 6 include Kituba, a compromise between various dialects of Kikongo, and Sango, developed between closely related Ubangian languages. Both have occasioned controversy as to whether or not they were pidgins or creoles, including claims that they are simply mixed rather than simplified. Arguments such as the above apply in both cases.

## 5. Row 3

Row Three also constitutes a ‘natural class’: creole languages. All are characterized by an extreme variety of reduction (*pace* objections to which I have responded elsewhere). But they vary according to degree of mixture.

5.1. *Box 7: Creoles without Substrate Features*

Box 7 represents cases in which reduction is stark but mixture is only lexical, such as Hawaiian Creole English. In terms of reduction, this creole significantly simplifies standard English morphosyntax. The creole is, in the end, a distinctly mesolectal one, constituting a continuum from standard English to the other end of a cline occupied by a dialect hardly as removed from English as, for example, Sranan. Nevertheless, the most distinct form of Hawaiian Creole English has (cf. Bickerton 1977) only variable occurrence of articles and plural marking, and no subject-auxiliary inversion. It also replaces English verbal morphology with invariant preposed particles such as *wen* for past, *stei* for imperfective, and *go* for future, combining in ways foreign to the standard English lexical sources for the items (*yu bin stei go* ‘you were going’). This trait does not parallel any of the substrate languages in any significant way (e.g., Chinese relies on preposed particles as well, but in functions in no way *significantly* congruent with Hawaiian Creole English’s). Overall, Hawaiian Creole English’s grammar is congruent enough with that of creoles like Tok Pisin that its native acquisition is considered an impediment to success in classrooms in which standard English is the vehicle.

Yet unlike most creoles, Hawaiian Creole English does not give evidence of being a hybrid language in terms of grammatical structure. It is not a compromise between English and its substrate languages Portuguese, Chinese, Japanese, Korean, Philippines languages, or Spanish. Chinese lent it no serialization (the few examples of this sort adduced by Bickerton [1984, 175] in early stages of the creole—albeit he presented them to argue for ‘bioprogram’ influence rather than substratal—are not attested afterward); Japanese lent it no verb-final constituent order, nor did Philippines languages lend it verb-initial order. Such calques occurred in the pidgin Englishes that preceded the emergence of the creole (Bickerton 1984, 174) but were eliminated when the language became a cross-ethnic lingua franca.

Siegel (2000a) has, in fact, adduced certain grammatical parallels between Hawaiian Creole English and Chinese. However, these are a compact set, applying to highly particular constructions scattered throughout the grammar, and always entailing likeness of an imprecise degree. Hawaiian Creole English does not appear to be Chinese with English words even to the degree that Singapore English is, and certainly not to the degree that Haitian Creole is claimed to be Fongbe with French words by Lefebvre (1998), or even to the lesser degree that other creoles are documented to parallel their substrate languages. Rather, Hawaiian Creole English’s parallels with Chinese are sparse ones of a sort that only a diligent analyst such as Siegel could smoke out by being committed to finding them, rather like the Chinese influences on vernacular Malay varieties discussed in chapter 8 in section 9—and then, again, to a much lesser extent. Thus, properly speaking, Hawaiian Creole English in our typology shades from box 7 slightly into box 8, but only very slightly, analogously to Mandarin Chinese’s light Altaic influence nudging it slightly into box 5.

It appears, then, that the immense typological variation between the various source languages for Hawaiian Creole English impeded any single one from predominating significantly in terms of calquing. Unremarkably, the creole contains a wealth of lexical items from its substrate languages—but very little more. This cre-

ole shows that vast reduction can occur without significant mixture from other languages. Structurally, Hawaiian Creole English is not a hybrid of English and its substrate languages; it is basically, rather, an abbreviation of English structure, *tout court*. The controversy over whether its grammar reflects an innate bioprogram is notorious, but no scholar of the subject, other than Holm (1986), has argued that this grammar was significantly based on substrate calques, Holm having ventured that the creole took on West African traits from African sailors (an argument that would appear superseded by later research on the creole's origins by Sarah Roberts).

Another example is the creolized version of the Chinook Jargon pidgin. Created by speakers of a large number of languages from various families (Chinook, Kalapuya, Takelma, Shasta, Molala, Umpqua, and others), the creole had no structural feature specifically traceable to any of these languages, instead expanding along a generalized plan. (Pace Grant 1996, I do not consider features such as shortened clitic forms of pronouns or the development of definite articles as traceable to the substrate languages, given their cross-linguistic generality.)

### 5.2. Box 8: *Most Creoles*

On the other hand, box 8 represents most creole languages. These are languages starkly less complex than older languages on all structural levels, as I have argued elsewhere. Mixture, while mediated by a general tendency of simplification (cf. also Keesing 1988) extends to syntax. However, the substrate languages—Languages B—do not contribute phonology or morphology. Here would fall, for example, Haitian Creole. Haitian Creole is much less complexified than both French and West African languages like Fongbe, in terms of syntax, phonology, and morphology (McWhorter 2004). However, its creation by speakers of Fongbe is clear in its postposed determiner, clefting of predicates, verb serialization, and other features rendering it quite distinct from any nonstandard variety of French (RES = resumptive):

(6) (a) Fongbe

**Wá**    *ɔ*    **ɖɛ̀** Jan    **wá**    *ɔ*    víví            nú nò    tòn.  
arrive DEF RES John arrive DEF make.happy for mother GEN

(b) Haitian Creole

**Rive**    **a**    Jan    **rive**    **a**,    fé    manman li    kontan.  
arrive DEF John arrive DEF make mother 3S happy

‘The (very) fact that John arrived made his mother happy.’ (Lefebvre 2004, 118)

(7) (a) Fongbe

**É**    **só**    àsón **yì** àxì    mɛ̀.  
he take crab go market in

(b) Haitian

Li    **pran** crab    **ale** nan mache.  
he take crab go in market  
‘He brought the crab to the market.’ (ibid., 112)



But Haitian does not have specific inheritances from Fongbe phonology such as labiovelarized phonemes [gb] and [kp]. Also, its morphology is not modeled significantly on Fongbe's.

Lefebvre (1998, 318–20) argues otherwise, showing parallels between Haitian and Fongbe derivational affixes. But, in fact, there are two cases where Haitian has an affix that Fongbe lacks, one where Haitian lacks one that Fongbe has, and one where Haitian has a prefix where Fongbe has a suffix. Then, although both languages have nominalizing machinery, Fongbe does this with reduplication of a verb while Haitian does so with a suffix, and, meanwhile, there are other grammatical functions of reduplication in Fongbe that Haitian lacks. Overall, I do not see a significant parallel between Haitian and Fongbe in derivational morphology: that both languages have affixes marking cross-linguistically commonplace distinctions such as agentive, attributive, inersive, ordinal, and so forth does not qualify as the idiosyncratic type of parallel that ideally founds an argument for “relexification.”

Many creoles in box 8 are actually continua of dialects shading toward the standard variety of the lexifier, such as many English-based creoles of the Caribbean. In these cases, the mesolectal varieties would extend upward into box 5, qualifying as semicreole levels.

### 5.3. *Deep Creoles*

Finally, box 9 represents those creoles in which general reduction is as robust as in Hawaiian Creole English and other creoles, but mixture from substrate languages extends to morphology and/or phonology. Saramaccan is typical: there is the same kind of syntactic mixture as in creoles like Haitian, but its main substrate language Fongbe also contributes reduplication used to connote the resultative:

- (8) (a) Saramaccan  
       dí jabíjabí mánda  
       ‘the opened basket’
- (b) Fongbe  
       xàsùn hùnhùn 5  
       basket open.open DEF  
       ‘the open basket’ (ibid., 353)

Saramaccan, unlike Haitian, does retain Fongbe labiovelar stops [kp] and [gb]. It is also tonal like Fongbe, extending to some lexical pairs (*kaí* “to fall,” *kái* “to call”), and also inherits a version of Fongbe’s tone sandhi patterns, as described in chapter 2. Sranan, Saramaccan’s sister creole, would belong in box 8, however, because in it, Fongbe influence does not extend to grammatical reduplication, tone, and sandhi, or the typically West African labiovelar segments.

Tok Pisin similarly reflects pronominal patterns of Australian and Aboriginal languages in affixal distinctions of inclusive/exclusive in the first and second persons (inclusive *yumi* and exclusive *mitupela*) and dual (*yumitupela*), and its predicate marker calqued on subject-marking prefixes (*em i go* “he goes”). Its phonology includes prenasalized stops (*ndai* “die”) and various allophones unlike English’s and

modeled on substrate ones, such as fricative allophones of some stops (Wurm 1977, 513). Berbice Creole Dutch contains some inflectional affixes (as well as grammatical items) from its main substrate language Eastern Ijo, such as the plural marker *-apu* and a perfective marker *-te*.

## 6. Closer Up

Obviously, one could compose a finer-grained chart accounting for gradient degrees of mixture and reduction that the nine-box schema does not capture.

For example, Mandarin would extend somewhat into box 5, in that it does display a slight degree of structural influence from Altaic, as Hashimoto (1986) argues. However, the degree is moderate enough that I have chosen to place it in box 4; what distinguishes Mandarin from its sister Chinese languages is overwhelmingly a matter of degree of complexity. The Altaic inheritance is slight, and in no way renders it structurally peculiar within its family.

In the meantime, the Celtic transfer in English, if eventually more generally accepted by historical linguists, would leave English, too, straddling boxes 4 and 5. This mixture is not as deep as the Chinese impact on Singapore English, but robust nonetheless. In that light, this would motivate placing Afrikaans between boxes 4 and 5 as well, given its lesser degree of mixture than Singapore English's.

Box 5 would also submit to finer gradation in terms of reduction: Reunionnais French is significantly more reduced from French than English and Afrikaans are from Proto-Germanic, as are the mesolectal varieties of English-based Caribbean creoles. However, then, the English creole mesolects have richer syntactic influence from African languages—that is, motivating entire books—than Reunionnais French has from Malagasy or Bantu (influences which have required painstaking identification in assorted articles). For example, if there were no boxes in the chart and the languages were simply distributed in a blank space according to the degree of mixture and reduction that they display, then one could configure a single box that would subdivide like so, with more prototypically “box 4” cases like Persian to the left of that box:

TABLE 9.2. Finer-Grained Version of the Box 4/Box 5 Area

	some syntactic mixture	heavy syntactic mixture
light reduction	Afrikaans English	Singapore English
medium reduction	Réunionnais French	Mesolectal English creoles

Note that this classification implies that “semicreole” Afrikaans and “regular language” English are in fact qualitatively equivalent in terms of contact history. I believe that this indeed represents the reality: both have light structural influence

from other languages (i.e., Celtic in the case of English), and, upon analysis, as I demonstrated in chapter 4, English reveals itself as, in fact, *more* reduced from Proto-Germanic than Afrikaans (and thus I place it *under* Afrikaans in the chart).

In fact, even this placement of English reflects my venturing, in the face of current orthodoxy in English historical linguistics, that certain Celtic features were transferred into English such as the *-ing* gerund and *do*-support. By my reading, the arguments from the renegade school of scholars arguing that English is essentially a Celticized Germanic language are likely to be tempered but by no means refuted entirely. However, at this writing, this school are preparing arguments for a much vaster amount of Celtic influence on English, and, if the “untempered” version of the argument gains acceptance, then English would belong in the upper-righthand box above with none other than Singapore English, although *above* it, in that it is less reduced from Proto-Germanic than Singapore English is:

TABLE 9.3. Hypothetical Finer-grained Version of the Box 4/Box 5 Area, 2050 AD

	some syntactic mixture	heavy syntactic mixture
light reduction	Afrikaans	English Singapore English
medium reduction	Réunionnais French	Mesolectal English creoles

Certainly, in the real world that we live in, English cannot help seeming somewhat peculiarly placed in such a typology, which implies that English is every bit as much a “semicreole” as Afrikaans. However, the ready plausibility of Afrikaans as a “semicreole” as against English is due, quite understandably, to the fact that Afrikaans emerged among brown-skinned peoples amid forced servant relationships in colonial circumstances in the middle centuries of the previous millennium, just as other creoles did. It is inevitable that sociohistorical realities will affect our perceptions in such fashion. But in the strict linguistic sense, the circumstances of the colonial South African context occasioned the same degree of structural reduction as the circumstances of England after the Scandinavian invasions—and, in fact, it occasioned a lesser degree of mixture.

## 7. Synchrony versus Sociology

In that light, the placement of languages in the chart is based on their synchronic character. Increasing degrees of reduction or mixture do not always correlate meaningfully with simple degree of contact between speakers of languages, such that the boxes do not each correspond to particular sociological situations. This is because of the assorted variables that determine the synchronic result besides intimacy of contact.

For example, in the second row, AAVE and Shaba Swahili emerged among large groups of laborers speaking several languages, and yet the synchrony of AAVE places it in box 4 while that of Shaba Swahili places it in box 6. Much of the difference here is due to the fact that Shaba Swahili was developed by speakers of Bantu languages with very similar grammars, facilitating the transfer of idiosyncratic features such as morphology.

The creoles in box 8 and box 9 all developed in plantation settings. Saramaccan has deeper influence from Fongbe than Sranan because it was developed in isolated maroon communities where European language influence was lesser than on the coastal plantations where Sranan developed. But Berbice Dutch is as deeply mixed as Saramaccan even though its creators were not maroons: it appears to have been largely created by speakers of one language, Eastern Ijo, which rendered transferred inflections comprehensible to the majority. Meanwhile, the deeper mixture in Tok Pisin (and its sister creoles Bislama and Solomon Islands Pijin) was due partly to the typological closeness of the Oceanic languages that were spoken by so many of their creators, and partly because the creoles have long been used alongside the substrate languages in perpetuity.

Looking vertically, Vikings simplified English (box 4) while the children who created Hawaiian Creole English simplified it much more—even though they were being schooled daily in standard English. They appear to have created the creole as a marker of identity alongside speaking a standard variety (S. J. Roberts 2000). It is unclear that Shaba Swahili's creators had appreciably greater access to Swahili than plantation slaves usually did when creating creoles; the typological closeness of Bantu languages appears, again, to have been the deciding factor.

Therefore, the typology I present must not be taken as evidence for an elegant—yet mistaken—demonstration that 1) grammars mix according to intimacy of contact, and 2) grammars reduce according to learners' access to the target. There are other factors that contribute to the outcomes. One more example is creolized Chinook Jargon. The Native Americans speaking various typologically disparate languages who created the creole did so on reservations, where they surely had ample "contact." But the language they created was not a dense conglomeration of syntax, morphology, and phonology from all of the languages like, say, Wutun in chapter 5 which mixes Mandarin, Tibetan, and Mongolian in this fashion. What these people created was instead a rendition of a pidginized Chinook in natural language form—because that pidgin, born of different sociological conditions, already existed in the context to be utilized. That is, my chart classifies results, not their causes. It is not intended as a sociolinguistic demonstration.

## 8. A Necessary Reification

Thus the goals of the chart are a synchronic taxonomization, revealing that a language can be deemed less complex than its sisters to a moderate rather than extreme degree, and that such a claim can be valid even when there is no evidence of pidginization in its history, and no evidence of anything but minimal and scat-

tered structural mixture from other languages. All of the boxes except box 4 correspond to phenomena with established names in linguistics: box 1 is *lexical borrowing*; box 2 is *Sprachbund effects*; while box 3 is “*intertwined*” languages and what is increasingly recognized as representing *language areas*. Boxes 5 and 6 contain *semicreoles*, while boxes 7 to 9 are true *creoles*, with boxes 7 and 8 containing *mesolectal* creoles where substrate influence is nonexistent (e.g., Hawaiian Creole English) or moderate (e.g., Haitian) and box 9 more basilectal “deeper” or “*radical*” ones (e.g., Saramaccan).

What remains, then, is a final reification, as it were, of the languages in box 4, which this book argues represent a regular kind of language contact result. I suggest that these be termed *Non-hybrid Conventionalized Second-language*, or NCSL, varieties.

## 9. Prediction over Description

I intend my thesis not as a mere descriptive account of five languages, but as a predictive claim. Namely, I first hypothesize that no language without significant non-native acquisition in its history will be especially simplified in comparison with its sisters. There will be variation in complexity within a minor range, such as Ura compared to Sye in Oceanic (Crowley 2000), or Spanish compared to French or Romanian, or Finnish, whose consonant gradations are not as complex as those in Estonian. But my hypothesis stipulates that grammars like Riau Indonesian and Saramaccan are not just unlikely, but impossible without extensive non-native acquisition.

Importantly, an indigenous language may be “simple” in many ways irrelevant to complexity as I have defined it. Pirahã of South America, for example, lacks numbers, colors, and clausal embedding, corresponding to a curiously unelaborated cultural perspective among its speakers in these regards. It also has a tiny phonemic inventory: just ten segments in women’s language. But the language has two tones, and is thoroughly inflected for categories including aspect, evidentiality, and nominalization (Everett 1986).

Second, I hypothesize that not just some, but all languages with significant non-native acquisition in their histories will be significantly less complex according to my metric than their sisters. That is, there will exist no such language that remains an “Icelandic” amid its closest relatives. Nurse (1997), for example, ventures that Swahili, used mostly non-natively for centuries, is what I term an NCSL among Bantu languages. Kusters (2003) makes a similar case referring to inflectional morphology for the Mainland Scandinavian languages, whose contrast in overall complexity with Icelandic is not as striking as English’s, but still worthy of investigation. Austronesianist John Wolff (May 2004, personal communication) has reported that Tagalog, the main lingua franca among the Philippines languages, is somewhat less morphologically complex than other Philippines languages, and that its sister Cebuano is moving in a similar direction as increasing numbers of non-natives acquire it. Algonquianist Richard Rhodes also reports (2001 personal communication) that Cree is the most user-friendly Algonquian language (although Cree’s quite ample complexity shows

especially well the relative nature of the concept), because of its frequent acquisition by non-native speakers throughout its history.

I would predict that a similar case would be Hindi in comparison to other Indo-Aryan languages, either in its standardized form or in widespread nonstandard varieties. I would also predict that the Mandinka-Bambara-Dyula dialect continuum in the Mande group within Niger-Congo would be less complex than other Mande languages, given its widespread use as a lingua franca since antiquity. Moreover, Wolof, as spoken today in Senegalese cities, would be a similar candidate, likely less complex than other West Atlantic languages like Serer (e.g., in Senegal in 2000, native Wolof speakers repeatedly told me that non-native speakers from the countryside, increasingly numerous in Dakar, tend to reduce the paradigm of noun class particles to a single general one).

There is a caveat here, however: languages widely learned non-natively but via formal teaching are exceptions, because the influence of pedagogy and the written standard discourage the conventionalization of abbreviated renditions of the language, marginalizing and stigmatizing these as “mistakes.” An example would be Gil’s (2001, 364–65) observation that the lingua franca Russian of Daghestan only slightly simplifies standard Russian and meanwhile incorporates complex features from Caucasian languages. In this region, however, standard Russian is the language of the schoolroom and is omnipresent in writing.

Thus the relevant cases are varieties that developed before widespread literacy had developed anywhere in the world, or nonstandard varieties that develop below the radar screen alongside official standard ones, such as Shaba Swahili. The former cases will tend to have developed in antiquity; the latter cases are the usual situation today. For example, only a minority of Fula speakers control its most elaborated variety, with its massively elaborate battery of noun classes and attendant consonant mutations. Most speak it as a second (third or fourth) language, in a much less complex rendition (LaCroix 1967) that exemplifies the kind of process I describe in this book, despite its nonstandard status.

## 10. What about the Romance Languages?

In this book I have on the one hand often noted that the inflectional loss from Latin to Romance is not “normal,” while on the other hand elsewhere stressed that the loss was compensated for by ample complexification, as in chapter 1, section 3.4, and section 3.2 in this chapter in reference to Romanian. The reader may ask whether I consider the Romance languages to be NCSLs or not. The answer is yes—but to an extent too moderate to have been useful in this book.

Substratal influence on the Romance languages other than the incontrovertible case of Romanian has remained extremely controversial for a very long time—that is, no features have been identified that trace unequivocally to external influence (cf. the possible Frankish influence on French as discussed in Thomason and Kaufman [1988, 126–28]). Thus it is unclear whether most Romance languages would belong in the first column of the table (no mixture beyond the lexical) or the second

(syntactic mixture). But in terms of the opposite axis of reduction, evidence strongly suggests that whichever column they belong in, that in a finer-grained table, they would belong in a row in between the first (no reduction) and the second (moderate but robust reduction as in English, Black English, Réunionnais French, or Afrikaans). According to my predictive framework, the case loss in the Latin noun phrase alone demands explanation, given the retention of nominal case morphology in so many other Indo-European languages.

Specifically, I concur with De Dardel and Wüest (1993) and Goyette (2000) that the Romance languages trace to a single simplification of Latin in Italy, later diffused across Europe to take particular forms in various locations, in the same fashion as a single Arabic koine was likely at the root of the New Arabics. De Dardel and Wüest note that if the Romance languages were each separate mutations of Latin, then we would not expect them all to have recruited the same prepositions to replace case markers, such as *ad* for dative and *de* for genitive (contrasting with English using *to* and *of* in these functions while Dutch uses etymologically unrelated *aan* and *van*). Goyette (*ibid.*, 118–19) adds that we would also not expect all to retain nominal forms derived from the accusative, given that retention of the nominative is equally common after case loss, as in Welsh and Bulgarian. Based on my predictions in the above section, I would nothing less than expect that if Latin was imposed as a second language on populations across the Roman Empire, then it would be manifested in these locales in simplified form.

But in comparison to the contrasts I have shown in the case studies, the simplification in Romance was distinctly moderate. While the loss of case marking is obvious, plural marking remained (in contrast to its virtual eclipse in Persian), as did grammatical gender marking and modifier concord, the latter two riddled with irregularities in all of the daughter languages. And loss of verb morphology left behind ample numbers of paradigms, variant according to verb classes, as well as vast numbers of irregular verbs. Then there was the development of articles, the conditional-marking paradigm, clitics contravening general word order, and other features.

This ratio of loss to gain appears at first glance comparable to that in the New Arabics, in which, for example, a degree of eclipse of the verb measures and/or allomorphy was counterbalanced by the emergence of new tense and aspect particles, an inalienable possessive distinction, an indefinite article, and so forth. But, in fact, the Romance languages not only compensate for Latin overspecification but also for structural elaboration, and not just in features traceable to the single Early Romance ancestor, but in individual developments, to a degree lacking in the New Arabics.

For example, in Italian, the definite article is indexed not only to gender but also to phonology, with the masculine article occurring usually as *il* (singular) and *i* (plural), but as *lo* (singular) and *gli* (plural) before [f], [ts] and [s] + consonant (*il* *uomo* “the man,” *lo* *stato* “the state”).

In Portuguese, third-person direct object enclitics *o*, *a*, *os*, and *as* vary phonetically according to when the preceding verb form ends in *r*, *s*, or *a*, in which case they take an initial *l* while *r* and *s* (but not *a*) elide, while if the form ends in a nasalized vowel then the clitic takes an initial *n* (Spencer 1991, 363):

TABLE 9.4. Third-Person Clitic Allomorphy in Portuguese, with *Levar* “To Raise”

1S	levo	levo-o
2S	levas	leva-lo
3S	leva	leva-lo
1P	levamos	levamo-lo
2P	levais	levai-lo
3P	levam	levam-no
infinitive	levar	levá-lo

When indirect-object clitics precede direct-object clitics, assorted unpredictable portmanteau morphemes or complexes result (ibid., 364):

TABLE 9.5. Indirect and Third-Person Direct-object Clitic Combinations in Portuguese

	indirect	indirect + direct
1S	me	mo, ma, mos, mas
2S	te	to, ta, tos, tas
3S	lhe	lho, lha, lhos, lhas
1P	nos	no-lo, no-la, no-los, no-las
2P	vos	vo-lo, vo-la, vo-los, vo-las
3P	lhes	lho, lha, lhos, lhas

In other Romance languages, the phonetic distance between the underlying form of such clitics and their surface manifestation in combination increases, such as in Friulian with first-person singular indirect-object clitics and the third-person direct ones (Faggin 1997, 109):

TABLE 9.6. Indirect and Direct-object Clitic Combinations in Friulian

3S masc.	me lo > m'al
3S fem.	me la > m'e
3P masc.	me li > m'ai
3P fem.	me le > m'es

French has suppletive oblique clitics beyond the accusative, genitive *en*, and dative *y*; for example, *J'en prends deux* “I’ll take two of them,” *Je pense à ma mère* “I am thinking about my mother,” *J’y pense* “I am thinking about her.” While in New



Arabic, liaison is a morphophonemic rule applying solely to feminine “hidden *t*,” in French liaison is phonological, such that all final consonants appear before a vowel: *il faut* [il fo]; *il faut y aller* [ifotiale].

The New Arabics do not exhibit this degree of individual innovation. Instead, they largely exhibit common innovations traceable to the koine that they all evolved from. Crucially, examining the new Arabics in sequence, one does not encounter assorted individual overspecifications and structural elaborations to anything approaching the *degree* to which such developments are typical among the Romance languages. Rather, the New Arabics vary according to how much they reduce the language of the Qurʾān, and in their idiosyncratic variations upon a common rootstuff, such as choice of morpheme to encode a particular lexical or grammatical concept. Where they exhibit a post-Classical Arabic complexity, all or most dialects share it. No New Arabic has evolved a third gender as Romanian has innovated a new neuter. No New Arabic has developed new verb Measures other than the anomalously conservative ones, the Bedouin varieties. No New Arabic has developed new present and perfect paradigms as Aramaic has. And it is significant that just as Romance developed a new conditional marking paradigm, so has South Arabian—but not the new Arabics.

It would seem that the Arabic koine was a more extreme reduction of Classical Arabic than Early Romance was a reduction of Latin. This meant that Early Romance had richer resources than New Arabic, serving as a springboard for developments such as those described above. The new Arabics started with less, and thus have yet to wend into the individual idiosyncrasies that the Romance languages have. That is, for example, with Classical Arabic case markers eliminated completely, no variety has remodeled vestigial ones into new functions or developed brand new ones. My interest in this book has been in cases of reduction of this extent, significantly more extreme than the Romance one.

Finally, I have refrained from including Romance because there is no contrasting daughter of Latin that has developed in consistent isolation without being imposed on speakers of other languages, or a surviving language from the higher Italic level of which Latin was but one member, to serve as an example of how an Italic language would survive if not widely acquired non-natively.<sup>1</sup> That is, we have no Romance or Italic equivalent to Aramaic or South Arabian.

As such, an argument for simplification in Romance must appeal to other Indo-European subfamilies such as Slavic for a comparandum. This is the strategy of, for example, Goyette (2000). I concur with his conclusions, but have opted not to include a Romance chapter because the languages’ contrast with their ancestor is too moderate for the argumentational power I have sought in this book.

## 11. Note to the Creolist

In various treatments, creolists have objected to my work identifying creoles as less complex than older languages according to the fact that I treat creoles as having developed “abnormally.”

I hope to have shown in this book that, on the contrary, I hardly consider creoles alone among the world’s languages in exemplifying an unusual degree of simplifica-

tion. Rather, I see the same kind of simplification as having occurred in grand old standardized vehicles of great literatures. Creoles are simply an extreme of this simplification, although even here, at the end of a continuum.

It would be possible, for example, to fashion a version of table 9.1 with a more fine-grained tabulation of degree of reduction. French creoles would be classed as less reduced than Sranan, in retaining most of French's derivational morphemes in contrast to Sranan arguably retaining none (if we treat agentive *-man* as a universally likely recruitment accidentally paralleling English's *-man*, with Sranan's lack of lexically opaque *-er* more germane). Hawaiian Creole English, as a relatively mesolectal creole in which morphology is only variably elided, would qualify as a less reduced version of its lexifier than Chinook Jargon Creole, in which Chinook's fearsome battery of morphology is absent.

As I have mentioned, I consider this continuum nature to also exist between creoles and other languages, most vividly illustrated by the creoles that themselves exist as unbroken continua between acrolectal standard varieties and basilectal "true creole" varieties.

Given my years of work on creoles, their contrast in simplification with older languages happens to have struck me first. I first became interested in creoles one blowsy night in Palo Alto, California, early in my graduate career, when, reading through Derek Bickerton's *The Roots of Language*, I came across data from São Tomense Creole Portuguese. I was intrigued by the fact that the creole was obviously much simpler than any Romance language, to the point that it seemed indisputable that some disruption in transmission had taken place as Bickerton argued, and yet that this was also clearly a natural human language like any other (as was clear from Bickerton's careful attempt to come to conclusions about its tense and aspect marking despite the system's complexity having challenged various analysts).

Over the years I found that most creolists found the hybrid nature of such creoles more interesting than their apparent relative simplicity compared to older languages. However, during this same period I learned that there is hybridity in, basically, all languages, and in many cases, much more than in creoles (e.g., Media Lengua, or areal phenomena in Australia, the Amazon, or the "Sinosphere"). I was drawn to investigate just what distinguished creoles besides the hybridity. The simplicity issue gradually became a focus of my interest after my earlier work investigating the African contribution to such languages.

As such, the basic issue of why any grammar simplifies radically instead of remaining like Hausa and Korean has always been my guiding commitment. This book represents my interest in this vein broadening to language contact theory in general. As such, I am committed not to a curious "Neo-Darwinian" (in DeGraff's terminology) quest to balkanize creoles as linguistic freaks of nature. Nor am I under an impression, explicit or tacit, that people of African descent are somehow incapable of acquiring new languages well—which would be a curious claim on my part given my book-length argument in *The Missing Spanish Creoles* that slaves on Spanish plantations regularly learned full Spanish under conditions that most creolists insist would have "filtered" language input and created creoles.

Rather, I proceed from a fundamental assumption that under typical conditions, human language remains always as complex as an Algonquian language, and I am

moved by an interest in investigating whether where this is not true—in not only Saramaccan and Haitian but also in my own native language, which I have argued above may well be a “semicreole”—it can be traced in systematic fashion to sociohistorical interventions. This involves not only contrasting Saramaccan and Haitian with Fongbe but also English with Icelandic, Malay with *Tukang Besi*, Moroccan Arabic with Aramaic, Mandarin with Cantonese, Persian with Pashto. I assume that the length alone of this book suggests that these contrasts between older languages are as compelling to me as those between them and creoles.

However, I openly assert that creoles are the product of a process of language transmission that is most definitely *abnormal*. I designate creoles’ development as abnormal because the sociohistorical nature of their timeline is much less common than the timeline of thousands of other languages worldwide. That is, their development was not *the norm*. However, this book has been devoted to arguing that the development of many noncreole languages, including the one I am writing in which is my native language, was also abnormal. The development of both English and Haitian Creole was abnormal—and fascinatingly so.<sup>2</sup>

## 12. Conclusion: Object Lesson

I will close with an example of a report that my typology provides a place for. Kotsinas (2001) describes a type of Swedish spoken by children of immigrants, which simplifies standard Swedish to a degree that naturally elicits remark as exceptional, but is in no sense the vastly reduced pidgin Swedish that the speakers’ parents tend to speak.

In agreement marking on determiners and adjectives, there is a tendency to replace the more marked neuter gender with common gender. General prepositions such as *på* and *i* tend to replace more marked ones. Fine-grained lexical semantic distinctions can be eroded, such as the typical Germanic one between *gå* “to walk” and *åka* “to travel by vehicle” (pointedly, absent in English). Thus these speakers will say *Jag ska gå till Grekland i sommar* “I will go to Greece this summer” rather than using *åka* as standard Swedish requires. Kotsinas also refers (without examples) to “word order deviations,” presumably flouting the structural elaboration of V2 rules that standard Swedish—but not English—retains as part of the Germanic heritage.

Kotsinas specifies that these traits do not trace to any of the home languages of the speakers—“it is impossible for an adult Swede to guess the so-called home language of individual adolescents” (ibid., 151)—despite the fact that the speakers still use their parents’ languages at home. The new variety does not have Turkish’s or Romany’s SOV word order, overgeneralized ablaut marking of the past tense in strong verbs modeled on Arabic’s nonconcatenational verb morphology, or Finnish’s absence of articles. Nor do these immigrant languages affect phonology in any significant way: there are certain intonational features used by all speakers, but none tracing specifically to Greek, Turkish, Finnish, Romany, or Arabic.

Rather, there is a new lingua franca that contrasts with Swedish most in a moderate degree of structural simplification, period—rather than in novel additions or remodelings based on the grammars that the speakers’ parents speak natively and

passed on to their children in the home. Otherwise, there are, unremarkably, ample lexical borrowings from the immigrant languages.

But Kotsinas, influenced by the tendency common among linguists to associate any degree of structural reduction with pidgins and creoles, terms this Swedish variety a semicreole, implying that it belongs in box 5 with Réunionnais, Afrikaans, and Singapore English. The problem here is that the variety she describes hardly evidences the degree of starkly foreign mixtures in semicreoles, such as the consecutive and serial constructions in Réunionnais, the double negative and genitive markers in Afrikaans, or the topic-comment structure and serialization in Colloquial Singapore English. On the impressionistic level, Kotsinas's account leaves one thinking that what she describes does not seem nearly as divergent from standard Swedish as Réunionnais French is from Parisian, Afrikaans is from standard or even regional Dutch, or Singapore English from standard English. Crucially, she likens the difference to that between AAVE and standard English.

However, she assumes that AAVE is a semicreole based on the tradition that places that dialect in box 5, whereas I have placed it in box 4 for the reasons adduced in 4.2.2. Similarly, the Swedish of second-generation immigrants that Kotsinas describes classifies as an NCSL variety. Complexity is reduced to a degree moderate but unparalleled in older nonstandard Swedishes, in which there are cases that retain more rather than less of the Proto-Germanic legacy, such as varieties like Alvdälska Swedish that retain three genders. Crucially, there is no evidence of grammatical mixture, and it would also be implausible to suppose that these speakers' Swedish has been reconstituted from the pidgin Swedish of the parents—in which case we would expect a Swedish creole grammatically parallel to Hawaiian Creole English or even Haitian Creole French, with preverbal particles replacing tense and aspect affixes and their allomorphs, and so on.

Rather, these speakers' language is Swedish slightly abbreviated but not mixed with other grammars beyond lexical items. This is not semicreole or "creoloid" Swedish, nor does it indicate that these speakers ever spoke the pidgin variety of their parents. It is a present-day manifestation of the kind of transformation that English underwent after the Scandinavian invasions.

Today, widespread literacy and education, and the prescriptive pressure of standard Swedish, will render this second-generation variety of Swedish a passing fashion. Even if it lives on, this will only be as an oral variety always considered an upstart departure from the "real" Swedish emblazoned in writing and taught in schools. But there was a time when language use among humans was predominantly oral, while written language was an elite peculiarity that waxed and waned over time. As such, at the periodic rebirths of this elite peculiarity after a period of eclipse, earlier written norms were often forgotten or discarded, such that a new standard could develop based on the oral reality of the language as it had changed since the decline of the old written standard. Hence, for example, Middle English after centuries of French rule, bringing to the page the non-native rendition of Scandinavians that had rendered Old English a new language, or Middle Persian under the Sassanid regime, bringing into writing what had happened to spoken Old Persian under the Achaemenids.

What distinguished these varieties was that between the decline of the old standard and the officialization of the new one, contact had decisively transformed the languages, in impeding their natural retention of needless complexity. In the case of the New Arabics, writing largely captures the grammars in the works of linguists and anthropologists; for the most part, an ancient written standard has persisted alongside evolving oral realities, with historical circumstance never occasioning new written standards reflecting those realities. But the New Arabics, though largely unwritten, contrast with Classical Arabic in the same fashion as the other languages contrast with the antique written stages of their heritage.

These new versions of the languages were not, nor had they ever been, pidgins or creoles. Nor, however, were they products of ordinary processes of local simplification—which themselves yield Polish, not Persian or Papiamentu. Nor were they simply language hybrids, given that centuries later, linguists would be frustrated in attempts to find in them calques upon the grammars that their creators had spoken natively. They were, instead, NCSL varieties, the occupiers of box 4 in my schematic typology.

To return to the opening of this book, the concept elicits little objection when observed in passing. “A lingua franca needs to be easy to grasp, and Malay has a more approachable structure than its relatives,” writes Dalby (1998, 391). Hopefully, at this point linguists will find it equally unobjectionable to apply the basic assumption in Dalby’s observation to sustained and predictive analyses of languages around the world.

# Notes

## Chapter 1

1. Polysynthetic languages, generally spoken by small groups, allow a considerable amount of null anaphora and have no distinct subordination construction. But as Hill (1993, 451–52) argues, there is no correlation between societal structure and this typology. The Aztecs created a hierarchical civilization, and yet Nahuatl is polysynthetic; meanwhile, languages of southern Australia are not polysynthetic, despite their speakers' having been small bands of hunter-gatherers. Moreover, polysynthetic languages remain replete with the features Perkins (1992) refers to, such as allomorphy, suppletions, and fine semantic distinctions.

## Chapter 2

Thank you to Reeli Torn for advising me on my presentation of the Estonian data, and for sharing her language.

1. Technically, a relatively compact phonemic inventory may entail so many allophonic distinctions overall as to encompass a larger number of possible segments than a relatively large phonemic inventory where overall allophonic variation is relatively minimal. However, these are possibilities, not necessities; it is not true that this kind of balancing out *must* be the case. For example, I am aware of no Polynesian language in which allophony renders the overall phonetic system as complex as Tsez's.

2. In the Estonian linguistic literature, alternations of length (*lipp:lipu*) are classified as quantitative; alternations of place and manner of articulation and voicing are classified as qualitative.

## Chapter 4

1. Page numbers refer to the following sources: Dutch: Donaldson (1997); Swedish: Holmes and Hinchliffe (1997); Danish: Thomas (1991); Norwegian: Strandskogen and Strandskogen (1986); Afrikaans: Ponelis (1993); Frisian: Tiersma (1985); Icelandic: Kress (1982); Faroese: Lockwood (1955); Yiddish: Lockwood (1995). Throughout the paper, where not cited, the source of the data is these (e.g., in cases of negative evidence such as the absence of a feature in a grammar, which is often impossible to refer to by page).

2. The construction is recessive in standard Dutch; speakers accept *Men heeft zijn arm gebroken* as well.

3. Shortly after writing this I noticed that *thence* is used occasionally in nonfiction prose. However, the English *to*-forms *hither* and *thither* are definitely impossible in Modern English beyond the ironic or deliberately archaic.

4. Pages for data in sources listed in note 1: Dutch: 125, Yiddish: 59, Swedish: 115–16, Icelandic: 97, Faroese: 57–61; Afrikaans data from Eksteen (1997); Frisian data in this table from Peter Tiersma (July 2001, personal communication).

5. The Mainland Scandinavian varieties differ in the fashions and extents to which they indicate direction in their adverbs, but the distinction is overall very much alive in both Danish and Norwegian.

6. Rydén and Brorström (1987, 211) also include *change*, *recover*, *turn* (in its transformative meaning), *set* (as in *The sun is set*), and *fly*. However, this author's intuitions as a native American English speaker sense these as strictly archaic, and suggest that *go* is indeed the only remaining form in the modern language, in concurrence with authors like Christophersen and Sandved (1969, 221).

7. This refers to the separation of auxiliary and main verb by objects, adverbials, etc. (*Ich habe den Brief gestern geschrieben* "I wrote the book yesterday").

8. The term includes Low German, which some contemporary observers suggest was processed as "the same language" at the time (Peters 1987, 80).

9. If by chance any of these attestations of [i] were actually remnants of the southwestern -y intransitive marker (*Idle chap, He'll do nothèn but fishy* [W. Barnes 1886, 25]), then even this—the preservation of an overt valence marker alien to English dialects northward—supports the idea that contact with Old Norse left English shorn of overspecification beyond what would have been the case amid internal evolution.

10. In the Shetlands and the Orkneys, where Norse was spoken for centuries longer than in England, the *be*-perfect has been generalized to transitives (Melchers 1992) rather than contracting and disappearing. However, this cannot be taken as counterevidence that Scandinavian contact spurred the demise of the *be*-perfect in England. As Melchers and myself note, the *be*-perfect was already quite restricted in Old Norse, such that the developments in the Shetlands and Orkneys cannot be seen as a transfer. Melchers (608) suggests that the culprit may have been southwestern Norwegian dialects where the *be*-perfect has extended to the transitive in just this way, given that most settlers of the Shetlands and the Orkneys came from southwestern Norway.

11. Original German: "Die Klagen soll man zuerst aburteilen, über die man im vorigen Sommer nicht zum Urteil gelangte."

## Chapter 5

1. In this chapter, I have opted to suspend for the purposes of this book the Siniticist tradition of referring to historical periods by dynasty alone. Because the sequence and timing of the dynasties is not common knowledge beyond those specializing in issues relating to China, this practice often interferes with ready engagement with the diachronic literature on Chinese for outsiders to the subfield. Given that this book is aimed at a general linguistic audience, I have chosen to "translate" the dynastic periods into more conventional century-based designations.

2. I have chosen to base my analysis upon a conventional seven-dialect classification rather than a more fine-grained one distinguishing, for example, the Jin, Hui, and Pinghua dialects (e.g., Chappell 2001a), for the mundane reason that research on the latter groups is

so sparse. However, indications known to me strongly suggest that this will not significantly impact my argumentation.

3. In Toishan Cantonese, tone alone can encode the perfective and the plural (Chappell 2001a, 331; Matthews and Yip 1994, 26). This, however, qualifies not as an overspecification but as a mere alternate strategy. Neither can it be seen as a structural elaboration, given that grammatical tone in itself cannot be seen as more “complex” than segmental indication of grammatical functions.

4. That Cantonese (and other Southern dialects) also uses classifiers optionally in possessive and relative clause constructions does not qualify as a complexification. To the extent that these usages, being optional, add no semantic distinction in themselves, they represent a kind of free variation which, if potentially analyzable as “complex” in this, would fall outside of the particular metrics I prefer to found this monograph upon. Meanwhile, where these strategies represent distinctions in register conditioned by the prestige of Mandarin, I see little usefulness in assessing relative complexity here given that Mandarin has its register layers as well.

5. Might I depart from the tonologist tradition of giving tone letter symbols and the corresponding numerical notation without explanation for novices? The numbers represent height of pitch with 5 the highest and 1 the lowest. The two- or three-figure numbers represent contour tones, such that 53 signifies a tone that starts at the highest pitch and falls to one in the middle; similarly, 213 signifies one that starts rather low, dips to the bottom and then rises to the middle.

6. Yet it is also true that about a third of Mandarin’s list are either synonym doublets (*dǎ-cóng* “from-from”) or aspect-marked versions of items that otherwise occur free, a remnant of the source of these “prepositions” in verbs (*wàng-zhe* “toward”). Neither of these constructions increases overspecification. But a substantial difference between Mandarin and Min in prepositional equipment remains.

7. While Lee-Smith surmises that Hezhou emerged when an Altaic variety was relexified by Mandarin rather than vice versa, linguistic and sociohistorical data on this and Tangwang are sparse, and at the state of our knowledge, it is unclear upon what linguistic or sociohistorical grounds we can define Hezhou as originally Altaic but Tangwang as originally Mandarin. The data would appear to readily allow that both varieties began as Mandarin. Varieties such as the Linxia Mandarin dialect (discussed below) can be seen as supporting this reconstruction, since they are indisputably Mandarin but have Altaic and Altaic-modeled morphology and syntax to a moderate degree. This suggests that both Hezhou and Tangwang represent an extreme along a pole of increasingly intense contact pressure upon Mandarin from Altaic languages, with dialects like Linxia falling in the middle.

8. Whether or not this feature is due to Altaic contact, the usual pattern is again demonstrated: Beijing Mandarin leans toward underspecification while its sisters lean in the other direction.

## Chapter 6

1. The distinction between counterfactual and conditional is my own, but appears to correspond clearly to MacKenzie’s (1990, 561) description of the semantics, that is, “‘(if) it were falling’ or ‘were to fall’” versus “‘(then) it would fall’.”

2. Sources for languages: Persian: Windfuhr 1979, 1987, Lazard 1992; Balochi: Barker and Mengal 1969, Elfenbein 1989; Kurdish: Khan and Lescot 1970, Joyce Blau 1989; Sivandi: Molchanova 1997; Farziandí: Rastorgueva and Moshkalo 1997; Gurani: Pireiko 1997;



Sangesari: Rastorgueva 1991a; Semnani: Pakhalina 1991; Zaza: Selcan 1998; Tati: Yar-Shater 1969; Gilaki: Rastorgueva et al. 1971; Talishi: Schulze 2000; Pashto: Trumpp 1873, MacKenzie 1987; Ormuri: Kieffer 2003; Ossetic: Thordarson 1989, Bagaev 1965; Yaghnobi: Khromov 1972, Bielmeyer 1989; Ishkashmi: Grierson 1920; Munji: Skjaervø 1989b, Oktor Skjaervø 2004, personal communication; Bartangi: Karamkhudoev 1973; Shugni: J. Payne 1989; Wakhi: Lorimer 1958, J. Payne 1989.

3. In this particular regard, Persian is overspecified compared to many of its sisters in having any animacy distinction here.

4. Strictly speaking, there was not a pathway from Middle to Modern Persian per se, as the former was not the direct ancestor of the latter. It is assumed that many early Iranian varieties are lost to history, never having been used in writing. Technically, it would appear that what was affected by the change was less a single variety of Persian than a range of closely related ones.

5. Sundermann instead supposes that the contrast indicates that it was Middle Persian that was especially interrupted in transmission, thereby requiring deliberate reconstitution in its written form. However, the motivation for this hypothesis is unclear. Even languages transmitted continuously have often been submitted to artificial codification, complete with decreed rules marginal or alien to actual speech. Classical Arabic is often thought to be such an example (cf. Kaye 1990, 665), and, in essence, all standardized languages also fall under the rubric.

## *Chapter 7*

1. Sources for data: Iraqi: Altoma 1969, Erwin 1969; Syrian: Cowell 1964, Grotzfeld 1965; Sudanese: Bergman 2002, Trimmingham 1946; Maltese: Borg and Azzopardi-Alexander 1997, Sutcliffe 1936; Moroccan: Caubet 1993; Gulf: Holes 1984, 1990; Egyptian: T. F. Mitchell 1956; Libyan: Owens 1984; Nigerian: Owens 1993. Brustad (2000) was also useful for supplementary data on Egyptian, Syrian, Gulf, and Moroccan. Najdi data is mainly from Ingham (1994) but as it is intended to represent Bedouin varieties as a whole, these data are confirmed as such, rather than being local only to Najdi or a minority of Bedouin dialects, in Rosenhouse (1984), from which source data on constructions not covered by Ingham are also taken (such as on the comparative [104–5]).

2. This Arabist notational style will not be used in this chapter, for the purposes of immediate clarity to other linguists. For the record, in the literature consonantal root patterns are traditionally designated via symbolizing the three consonants with [f], [ʕ], and [l]. Thus Measure Three's perfect form of *kataba* "he wrote," *kātaba*, is the *faaʕala*.

3. I mark the analytic comparative as marginal in Bedouin varieties based on the fact that although Grand'Henry documents it in the Algerian variety (1976, 66), two seminal treatments do not mention it (Ingham's [1994] Najdi grammar and Rosenhouse's [1984] description of the North Israel variety and cross-dialectal survey), while Cohen (1963, 212) explicitly writes that "all" triliteral adjectives occur as morphological comparatives in the Mauritanian variety. Given how readily grammarians note the analytic comparative in other new Arabics, it seems reasonable to treat negative evidence as indicating the absence of the feature in the case of the Bedouin grammars.

4. Hence my eschewal of the traditional term "Neoarabic" in favor of "the new Arabics." The term would seem to carry an implication that "Neoarabic" is in some sense one language with many variations. But given the gulf between, for example, Moroccan Arabic and Gulf Arabic, it would seem that discussing these language varieties under a unitary term such as "Neoarabic" would require us to unhesitatingly discuss French and Romanian as mere variations of "Neo-Latin."

5. Kusters's inflectional focus leads him to discount these (129) as mood marking in being neither obligatory nor used with both aspects, but my broader purposes justify including them.

## Chapter 8

1. I will use the term *Malay* as a cover term for the two standard languages of Malaysia (Malaysian) and Indonesia (Indonesian); the differences between the two are largely lexical (e.g., Prentice 1987, 916), and, in general, are of no import to the thesis of the chapter. However, my data on this language are drawn from sources on Indonesian, more copious, and useful to the modern linguist, than ones on Malaysian.

2. On my refraining from using Malayo-Chamic (Thurgood 1999) as the “hesion” of interest, see section 5.

3. The dative affixes happen to be recessive in *Tukang Besi*, but are robust in close relatives (Donohue 1999, 136).

4. Sources for language data unless otherwise indicated: Malay: Sneddon 1996; Minangkabau: Moussay 1981; Iban: Omar 1981; Javanese: Horne 1961, Suharno 1982; Sundanese: Coolsma 1904, Müller-Gotama 2001; Madurese: Stevens 1968, Davies 1999; Toba Batak: Nababan 1981; Daro Batak: Woollams 1996; Macassarese: Mathes 1858, Cense 1979; Buginese: Sirk 1979; Nias: Brown 2001; *Tukang Besi*: Donohue 1999; Muna: Van den Berg 1989.

5. This is not a structural elaboration unique to Malay, occurring also in, for example, its close relatives Sundanese and Madurese.

6. Note that the complexity of Kerinci Malay in its nouns with distinct roots sentence-medially versus sentence-finally is not a retention but an innovation. Unknown in other Malayic varieties, and thus occurring after Proto-Malayic, this trait emerged while, or after, the variety that became standard Malay was abbreviated by non-native acquisition and was then standardized. The question is why regional Malays do not always, regularly, or even commonly have robust innovations of this kind extending to entrenched and intricate features like allomorphy and suppletivity. The innovations in Kerinci Malay are extensive in the fashion that French's are in comparison to Latin, or Modern Aramaic's are compared to Classical Aramaic. The innovations in Riau Indonesian are nothing of the sort, and the qualitative difference demands, I believe, an explanation.

7. Hull (1998, 167–68) considers the languages of Timor to classify as Western Malayo-Polynesian, noting close lexical affinity with languages of southeast Sulawesi, and arguing that the proposed Central Malayo-Polynesian group is no more traceable to a proto-language than Western Malayo-Polynesian. By my reckoning, Hull's classification appears likely to prevail ultimately. However, my reckoning is not that of a career Austronesianist, and, as such, because at this writing Hull's analysis is not yet generally accepted, I opt to treat these languages under their traditional classification as members of the Central Malayo-Polynesian subfamily.

8. Hull (1998, 165; 2001, 101) supposes that contact with the PMD Malay of Ambon further encouraged analytic structure in Tetun Dili (and degrees of the tendency in other languages of Timor). But this leads to questions as to why so many Austronesian languages remain heartily synthetic despite contact with similar Malay varieties, and why other languages spoken alongside creoles for centuries have not become analytic (such as the West Atlantic languages spoken by users of Guinea-Bissau Creole Portuguese). I would suggest that the theoretical necessity of a second phase of simplification due to PMD Malay is unclear.

## Chapter 9

1. Romanian does have a genitive/dative case marker. But De Dardel and Wüest (1993) deftly attribute this retention to later prescriptive influence from Latin, arguing that only

Romanian and early Gallo-Romance diverge from the otherwise pan-Romance hallmark of retaining only accusative forms, and in fashions quite distinct from one another, suggesting developments that occurred after the initial emergence of a single Early Romance variety in Italy. However, the data also allow that Early Romance had case marking beyond the accusative, and that the difference in its retention and reduction between Romanian, early Gallo-Romance, and the other Romance languages was a matter of the same minor “flutter” in degree of reduction that distinguishes languages like Oceanic Ura and Sye today.

2. I will assume that the sentence “creoles are the product of a process of language transmission that is most definitely *abnormal*” will not be cited in isolation as a demonstration of dismissive attitudes toward creole languages, with an implication that the sentence did not occur within a careful exposition of a case for the claim, including the subsumption within it of languages like English.

## References

- Abas, Husen, and Charles E. Grimes. 1995. Bugis. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 550–61. Berlin: Mouton de Gruyter.
- Abun-Nasr, J. M. 1987. A history of the Maghrib in the Islamic period. Cambridge: Cambridge University Press.
- Adelaar, K. Alexander. 1992. Proto-Malayic: The reconstruction of its phonology and parts of its lexicon and morphology. Canberra: Research School of Pacific Studies (The Australian National University).
- . 1995. Minangkabau. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 433–41. Berlin: Mouton de Gruyter.
- Adelaar, K. Alexander, and D. J. Prentice (with C. D. Grijns, H. Steinhauer, and A. van Engelenhoven). 1996. Malay: Its history, role and spread. In *Atlas of languages of inter-cultural communication in the Pacific, Asia, and the Americas*, vol. 2:1, ed. Stephen. A. Wurm, Pieter Mühlhäusler, and Darrell T. Tryon, 673–93. Berlin: De Gruyter.
- Ahlgren, Arthur. 1946. On the use of the definite article with “nouns of possession” in English. Uppsala: Appelbergs Boktryckeriaktiebolag.
- Aikhenvald, Alexandra Y. 1999. Areal diffusion and language contact in the Içana-Vaupés basin, north-west Amazonia. In *The Amazonian languages*, ed. R. M. W. Dixon and Alexandra A. Aikhenvald, 385–416. Cambridge: Cambridge University Press.
- . 2001. Areal diffusion, genetic inheritance, and problems of subgrouping: A North Arawak case study. In *Areal diffusion and genetic inheritance*, ed. Alexandra Y. Aikhenvald and R. M. W. Dixon, *The Amazonian languages*, 167–94. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra Y., and R. M. W. Dixon. 2001. *Areal diffusion and genetic inheritance*. New York: Oxford University Press.
- Altoma, Salih J. 1969. *The problem of diglossia in Arabic: A comparative study of Classical and Iraqi Arabic*. Cambridge, Mass.: Center for Middle Eastern Studies.
- Anceaux, J. C. 1952. *The Wolio language*. The Hague: Martinus Nijhoff.
- Anceaux, J. C., and Charles E. Grimes. 1995. Wolio. In *Comparative Austronesian dictionary*, part 1, fascicle 1, 573–84. Berlin: Mouton de Gruyter.
- Ansaldi, Umberto, and Stephen Matthews. 1999. Creolization and the Hokkien substrate in Baba Malay. *Journal of Chinese Linguistics* 27:38–68.

- Arends, Jacques. 2001. Simple grammars, complex languages. *Linguistic Typology* 5 (2/3): 180–82.
- Arnold, Werner, and Peter Behnstedt. 1993. *Arabisch-Aramaische Sprachbeziehungen im Qalamūn (Syrien)*. Wiesbaden: Otto Harrassowitz.
- Bagaev, Nikolai Konstantinovich. 1965. *Sovremenniy osetinskii jazyk*. Ordzhonikidze: Severo-Osetinskoe.
- Bailey, Charles-James, and Karl Maroldt. 1977. The French lineage of English. In *Pidgins-creoles-languages in contact*, ed. Jürgen Meisel, 21–53. Tübingen:
- Narr. (In answer to (3): the capitalization of “pidgin” and “creole” varies massively according to author, school of thought, era, and so on. And then, the rules about capitalization in titles makes it all even messier, and never mind that Germans capitalize everything they can while the French avoid it like the plague, and so on—despite that the French are most likely to consider creoles to be an entity called “Le Creole.” I will adopt a house style of using lower case as a default, with caps only when the authors’ usage, etc. make it clear that they consider “pidgin” or “creole” to constitute an entity that gets a capital letter. This is often the case, e.g., when French writers refer to a particular creole as a language called “Creole,” in which case to not use the capital would be like writing “english.”)
- Baird, Louise. 2002. A grammar of Kéo: A language of East Nusantara. PhD diss., Australian National University.
- Baker, Philip. 1984. Agglutinated nominals in creole French: Their evolutionary significance. *Te Reo* 27:89–129.
- Baker, Phillip, and Chris Corne. 1982. *Isle de France Creole: Affinities and origins*. Ann Arbor, Mich.: Karoma.
- Bakker, Peter. 1997. *A language of our own*. New York: Oxford University Press.
- Bakker, Peter, and Pieter Muysken. 1995. Mixed languages and language intertwining. In *Pidgins and creoles: An introduction*, ed. Jacques Arends, Pieter Muysken, and Norval J. Smith, 41–52. Amsterdam: John Benjamins.
- Ball, M. J., and N. Müller. 1992. *Mutation in Welsh*. London: Routledge.
- Ballard, W. L. 1985. The linguistic history of South China: Miao-Yao and southern dialects. In *Linguistics of the Sino-Tibetan area: The state of the art*, ed. Graham Thurgood, James Matisoff, and David Bradley, 58–84. Canberra: Australian National University.
- Bao, Houxing, Ruoyun Shen, Zhenhua Cui, and Yunji Wu. 1999. *Changsha fangyan yanjiu* [Study of the Changsha dialect]. Changsha: Hunan Education Press.
- Bao, Zhiming. 2005. The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics* 41: 269–305.
- Baptista, Marlyse. 2000. Verb movement in four creole languages: A comparative analysis. In McWhorter, *The missing Spanish Creoles: Recovering the birth of plantation contact languages*, ed. John H. McWhorter, 1–33. Berkeley and Los Angeles: University of California Press.
- Barker, Muhammad Abd-al-Rahman, and Aqil Khan Mengal. 1969. *A course in Baluchi*, vol. 1. Montreal: McGill University Institute of Islamic Studies.
- Barnes, Dayle. 1977. To *er* or not to *er*. *Journal of Chinese Linguistics* 5: 211–36.
- Barnes, Janet. 1984. Evidentials in the Tuyuca verb. *International Journal of American Linguistics* 50:255–71.
- Barnes, William. 1886. *A glossary of the Dorset dialect with a grammar of its word shapening and wording*. London: Trübner.
- Barr, Donald F. 1995. Da’a. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 529–37. Berlin: Mouton de Gruyter.

- Bartholomae, Chr. 1883. *Handbuch der Altiranischen Dialekte*. Leipzig: Breitkopf & Härtel.
- Baxter, William H. 2000. Did Proto-Mandarin exist? *Journal of Chinese Linguistics* 28:100–115.
- Benjamin, Geoffrey. 1993. *Grammar and polity: The cultural and political background to Standard Malay. The role of theory in language description*, ed. William Foley, 341–92. Berlin: Mouton de Gruyter.
- Bense, J. F. 1939. *A dictionary of the Low-Dutch element in the English vocabulary*. The Hague: Martinus Nijhoff.
- Bentley, Delia, and Thórhallur Eythórsson. 2001. Alternation according to person in Italo-Romance. In *Historical linguistics 1999*, ed. Laurel Brinton, 63–74. Amsterdam: John Benjamins.
- Bergman, Elizabeth M. 2002. *Spoken Sudanese Arabic*. Springfield, Va.: Dunwoody.
- Bichlmeier, Harald, and Roland Marti. 2003. From PIE to OCS: ALG or MAC? Review of Comparative Slavic nominal morphology: A new synthesis, by Robert Orr. *Canadian Slavonic Papers* 3–4:469–81.
- Bickerton, Derek. 1977. Change and variation in Hawaiian English, vol. 2: Creole syntax. National Science Foundation final report.
- . 1984. The language bioprogram hypothesis. *The behavioral and brain sciences* 7:173–221.
- . 1995. *Language and human behavior*. Seattle: University of Washington Press.
- Bielmeier, Roland. 1989. Yaghnōbī. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 480–88. Wiesbaden: Ludwig Reichert.
- Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. In *Numeral types and changes worldwide*, ed. Jadranka Gvozdanovic, 113–85. Berlin: Mouton de Gruyter.
- Bivar, A. D. H. 1983. The political history of Iran under the Arsacids. The Cambridge history of Iran, vol. 3:1, ed. Ehsan Yarshater, 21–99. Cambridge: Cambridge University Press.
- Blanc, Haim. 1970. Dual and pseudo-dual in the Arabic dialects. *Language* 46:42–57.
- Blau, Joshua. 1965. *The emergence and linguistic background of Judaeo-Arabic: A study of the origins of Middle Arabic*. New York: Oxford University Press.
- . 1977. The beginnings of the Arabic diglossia: A study of the origins of Neoarabic. *Afroasiatic Linguistics* 4:4.
- . 1983. Vestiges of *tanwin* -un and the case ending -u as attested in Yemenite Judeo-Arabic texts from the seventeenth century. *Bulletin of the School of Oriental and African Studies, University of London*, 46:529–31.
- Blau, Joyce. 1989. Le kurde. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 327–35. Wiesbaden: Ludwig Reichert.
- Bloch, Ariel. 1971. Morphological doublets in Arabic dialects. *Journal of Semitic Studies* 16:53–73.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Taipei: Academia Sinica.
- . 2002. Notes on the history of “focus” in Austronesian languages. In *The history and typology of Western Austronesian voice systems*, ed. Fay Wouk and Malcolm Ross, 63–78. Canberra: Pacific Linguistics.
- Boas, Franz. 1938. *The mind of primitive man*. New York: Macmillan.
- Bobaljik, Jonathan, and Hoski Thráinsson. 1997. Two heads aren’t always better than one. *Syntax* 1:37–71.
- Borg, Albert, and Marie Azzopardi-Alexander. 1997. *Maltese*. London: Routledge.

- Bosman, D. B. 1962. *Oor die ontstaan van Afrikaans*. Amsterdam: Swets en Zeitlinger.
- Braunmüller, Kurt. 2002. Language contact during the Old Nordic period I: With the British Isles, Frisia and the Hanseatic League. In *The Nordic languages: An international handbook of the history of the Nordic Germanic languages*, ed. Oskar Bandle, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann, and Ulf Teleman, vol. 1, 1028–39. Berlin: Mouton De Gruyter.
- Breeze, Andrew. 1993. Celtic etymologies for Old English *cursung* “curse,” *gafeluc* “javelin,” *stær* “history,” *syrc* “coat of mail,” and Middle English *clog(ge)* “block, wooden shoe,” *cokkunge* “striving,” *tirven* “to flay,” *warroke* “hunchback.” *Notes and Queries* 40(3): 87–97.
- Brenner, Oscar. 1882. *Altnordisches Handbuch*. Leipzig: T. O. Weigel.
- Briant, Pierre. 1996. *Histoire de l'empire Perse*. Paris: Fayard.
- Bright, William. 1957. The Karok language. *University of California Publications in Linguistics*, vol. 13. Berkeley and Los Angeles: University of California Press.
- Brilioth, Börje. 1913. *A grammar of the dialect of Lorton (Cumberland)*. Oxford: Horace Hart.
- Brown, Lea. 2001. A grammar of Nias Selatan. PhD diss., University of Sydney.
- Brunner, Christopher J. 1977. *A syntax of Western Middle Iranian*. Delmar, N.Y.: Caravan.
- Brustad, Kristen E. 2000. *The syntax of spoken Arabic*. Washington, D.C.: Georgetown University Press.
- Burrow, T. 1955. *The Sanskrit language*. London: Faber & Faber.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Cahen, Claude. 1975. Tribes, cities and social organization. *Cambridge History of Iran*, vol. 4, ed. R. N. Frye, 305–28. Cambridge: Cambridge University Press.
- Callaway, Jr., Morgan. 1913. The infinitive in Anglo-Saxon. Washington: Carnegie Institution of Washington.
- Caubet, Dominique. 1993. *L'arabe marocain*, vol. 1. Paris: Éditions Peeters.
- Cavalli-Sforza, Luigi Luca, and Francesco Cavalli-Sforza. 1995. *The great human diasporas*. Cambridge, Mass.: Perseus.
- Cense, A. A. 1979. *Makassaara-Nederlands Woordenboek*. The Hague: Martinus Nijhoff.
- Cervazhe, Iliia. 1965. *Andiuri ena* [The Andi language]. Tbilisi: Mecniereba.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.
- Chappell, Hilary. 2001a. Language contact and areal diffusion in Sinitic languages. In *Areal diffusion and genetic inheritance*, ed. Alexandra Y. Aikhenvald and R. M. W. Dixon, 328–57. New York: Oxford University Press.
- . 2001b. *Sinitic grammar: Synchronic and diachronic perspectives*. Oxford: Oxford University Press.
- . 2001c. A typology of evidential markers in Sinitic languages. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 56–84. Oxford: Oxford University Press.
- . 2005. *A typology of Sinitic languages: Morphosyntax and grammar*. New York: Oxford University Press.
- Chappell, Hilary, and Christine Lamarre. 2004. *A grammar and lexicon of Hakka: Historical materials from the Basel Mission Library*. Collection des Cahiers de Linguistique — Asie Orientale. Paris: Ecoles des Hautes Etudes en Sciences Sociales.
- Chaudenson, Robert. 1979. *Les créoles français*. Évreux: Nathan.
- . 1992. *Des îles, des hommes, des langues*. Paris: L'Harmattan.
- Chen, Chung-Yu. 1983. A fifth tone in the Mandarin spoken in Singapore. *Journal of Chinese Linguistics* 11:92–119.

- Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4:109–50.
- Cheng, Robert L. 1977. Taiwanese question particles. *Journal of Chinese Linguistics* 5:153–85.
- Christophersen, Paul, and Arthur O. Sandved. 1969. *An advanced English grammar*. London: Macmillan.
- Clark, Ross. 1987. Austronesian languages. In *The world's major languages*, ed. Bernard Comrie, 901–12. New York: Oxford University Press.
- Clarke, Sandra. 1982. *North-West River (Sheshatshit) Montagnais*. Ottawa: National Museums of Canada.
- Cleasby, Richard, and Gudbrand Vigfusson. 1957. *An Icelandic-English dictionary*. Oxford: Clarendon.
- Cohen, David. 1963. *Le dialecte arabe Ḥassānīya de Mauritanie*. Paris: Klincksieck.
- . 1970. Koinè, langues communes et dialectes arabes. In *Etudes de linguistique sémitique et arabe*, ed. David Cohen, 105–25. The Hague: Mouton.
- Colledge, Malcolm A. R. 1967. *The Parthians*. New York: Frederick A. Praeger.
- Comrie, Bernard, ed. 1987. *The world's major languages*. New York: Oxford University Press.
- Comrie, Bernard, Maria Polinsky, and Ramazan Rajabov. 2000. *Tsezian languages* (manuscript).
- Cook, J. M. 1983. *The Persian empire*. New York: Schocken.
- . 1985. *The rise of the Achaemenids and establishment of their empire*. In *The Cambridge history of Iran*, vol. 2, ed. Ilya Gershevitch, 200–291. Cambridge: Cambridge University Press.
- Coolsma, S. 1904. *Soendaneesche spraakkunst*. Leiden: A. W. Sijthoff.
- Corne, Chris. 1999. From French to Creole. London: University of Westminster Press.
- Cornips, Leonie. 2002. Convergence and divergence in grammar. In *The convergence and divergence of dialects in contemporary societies*, ed. P. Auer, Frans Hinskens, and P. Kerswill. Cambridge: Cambridge University Press.
- Corriente, Federico. 1971. On the functional yield of some synthetic devices in Arabic and Semitic morphology. *Jewish Quarterly Review* 62:20–50.
- Cowan, William G. 1966. Two notes on Arabic dialectology. *Journal of the American Oriental Society* 86:416–18.
- Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic*. Washington, D.C.: Georgetown University Press.
- Crowley, Terry. 2000. Simplicity, complexity, emblematicity and grammatical change. In *Processes of language contact: Studies from Australia and the South Pacific*, ed. Jeff Siegel, 175–93. Quebec City: AGMV.
- Crystal, David. 1987. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Cumming, Susannaa. 1991a. *The case of Malay constituent order*. Berlin: Mouton de Gruyter.
- . 1991b. *Functional change: The case of Malay constituent order*. Berlin: Mouton de Gruyter.
- Curme, George O. 1952. *A grammar of the German language*. New York: Frederick Ungar.
- Dahl, Östen. 2004. The growth and maintenance of linguistic complexity. Amsterdam: John Benjamins.
- Dahlstrom, Amy. 1991. *Plains Cree morphosyntax*. New York: Garland.
- Dal, Ingerid. 1966. *Kurze Deutsche Syntax*. Tübingen: Max Niemeyer.
- Dalby, Andrew. 1998. *Dictionary of languages*. New York: Columbia University Press.
- Dalby, Michael T. 1979. Court politics in late T'ang times. In *The Cambridge history of China*, vol. 3, ed. Denis Twitchett and John K. Fairbank, 561–681. Cambridge: Cambridge University Press.



- Dalton-Puffer, Christiane. 1995. Middle English as a creole and its opposite: On the value of plausible speculation. In *Linguistic change under contact conditions*, ed. Jacek Fisiak, 35–50. Berlin: Mouton de Gruyter.
- . 1996. *The French influence on Middle English morphology: a corpus-based study of derivation*. Berlin: Mouton De Gruyter.
- Danchev, Andrei. 1997. The Middle English creolization hypothesis revisited. In *Studies in Middle English linguistics*, ed. Jacek Fisiak, 79–108. Berlin: Mouton de Gruyter.
- Daniel, Elton. 2001. *The history of Iran*. Westport, Conn.: Greenwood.
- Davidoff, Jules, Ian Davies, and Debi Roberson. 1999. Colour categories in a stone-age tribe. *Nature* 398:203–4.
- Davies, William. 1999. *Madurese*. Munich: Lincom Europa.
- Debevoise, Neilson C. 1938. *A political history of Parthia*. Chicago: University of Chicago Press.
- De Dardel, R., and J. Wüest. 1993. Les systèmes casuels du proto-roman: Les deux cycles de simplification. *Vox Romanica* 52:25–65.
- DeGraff, Michel, ed. 1999. *Language creation and language change*. Cambridge, Mass.: MIT Press.
- . 2001. On the origins of creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5:213–310.
- Deletant, Dennis, and Yvonne Alexandrescu. 1992. *Teach yourself Romanian*. Lincolnwood, Ill.: NTC.
- De Mello, Heliana R., Alan N. Baxter, John Holm, and William Meggeney. 1998. O português vernáculo do Brasil. In *América negra: Panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, ed. Matthias Perl and Armin Schwegler, 72–137. Frankfurt: Vervuert.
- De Miroschedji, Pierre. 1985. La fin du royaume d'Ans an et de Suse et la naissance de l'Empire perse. *Zeitschrift für Assyriologie und Vorderasiatische Archäologie* 75:265–306.
- Den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. In *On the formal syntax of the Westgermania: Papers from the Third Groningen Grammar Talks, Groningen, January 1981*, ed. Werner Abraham, 47–131. Amsterdam: John Benjamins.
- . 1986. Double negation and the genesis of Afrikaans. In *Substrata versus universals in creole genesis*, ed. Pieter Muysken and Norval J. Smith, 185–230. Amsterdam: John Benjamins.
- Denison, David. 1993. *English historical syntax: Verbal constructions*. London: Longman.
- De Rooij, Vincent A. 1997. Shaba Swahili: Partial creolization due to second language learning and substrate pressure. In *The structure and status of pidgins and creoles*, ed. Arthur K. Spears and Donald Winford, 309–39. Amsterdam: John Benjamins.
- Diakonoff, I. M. 1985. Media. In *The Cambridge history of Iran*, vol. 2, ed. Ilya Gershevitch, 36–148. Cambridge: Cambridge University Press.
- Diem, Werner. 1973. Die nabatäischen Inschriften und die Frage der Kasusflexion im Altarabischen. *Zeitschrift der deutschen morgenländischen Gesellschaft* 123:227–37.
- . 1978. Divergenz und Konvergenz im Arabischen. *Arabica* 25:128–47.
- . 1979. Studien der Frage des Substrats im Arabischen. *Der Islam* 56:12–80.
- Dimmendaal, Gerrit J. 2001. Areal diffusion versus genetic inheritance: An African perspective. In *Areal diffusion and genetic inheritance*, ed. Alexandra Y. Aikhenvald and R. M. W. Dixon, 358–92. New York: Oxford University Press.
- Dixon, R. M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- . 2001. The Australian linguistic area. In *Areal diffusion and genetic inheritance*, ed.

- Alexandra Y. Aikhenvald and R. M. W. Dixon, 64–104. New York: Oxford University Press.
- Djawanai, Stephanus, and Charles E. Grimes. 1995. Ngada. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 593–99. Berlin: Mouton de Gruyter.
- Dol, Philomena. 1999. *A Grammar of Maybrat*. PhD diss., University of Leiden.
- Domingue, Nicole Z. 1977. Middle English: Another creole? *Journal of Creole Studies* 1:89–100.
- Donaldson, Bruce. 1997. *Dutch: A comprehensive grammar*. London: Routledge.
- Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- Dryer, Matthew. 2003. Word order in Sino-Tibetan languages from a typological and geographical perspective. In *The Sino-Tibetan languages*, ed. Graham Thurgood and Randy J. LaPolla, 43–55. London: Routledge.
- Durie, Mark. 1985. *A grammar of Acehnese on the basis of a dialect of north Aceh*. Dordrecht, Holland: Foris.
- Dwyer, Arienne. 1992. Altaic elements in the Linxia dialect: Contact-induced change on the Yellow River plateau. *Journal of Chinese Linguistics* 20: 160–78.
- Dyen, I. 1965. A lexicostatistical classification of the Austronesian languages. *International Journal of American Linguistics*, Memoir 19. Baltimore: Waverly.
- Edwards, John. 1994. *Multilingualism*. London: Penguin.
- Efimov, B. A., ed. 1997. *Osnovy iranskoivo jazykoznanija: Novoiranskije jazyki: Severo-zapadnaja grupa*, vol. 2. Moscow: Vostočnaja literatura R. A. N.
- Eksteen, L. C. 1997. *Major dictionary/Groot woordeboek*. Cape Town: Pharos.
- Elfenbein, Josef. 1989. *Balōči*. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 350–62. Wiesbaden: Ludwig Reichert.
- Emmerick, Ronald E. 1989. Khotanese and Tumshuqese. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 204–29. Wiesbaden: Ludwig Reichert.
- . 1998. Languages of the world (“The Iranian languages”). In *Encyclopedia Britannica* (Macropedia). Chicago: Encyclopedia Britannica.
- Erwin, Wallace M. 1969. *A basic course in Iraqi Arabic*. Washington, D.C.: Georgetown University Press.
- Evans, Nicholas, Dunstan Brown, and Greville Corbett. 2002. The semantics of gender in Mayali: Partially parallel systems and formal implementation. *Language* 78:111–55.
- Everett, Daniel L. 1986. Piraha. In *Handbook of Amazonian languages*, vol. 1, ed. D. C. Derbyshire and G. K. Pullum, 200–325. Berlin: Mouton de Gruyter.
- Fabian, Johannes. 1986. *Language and colonial power: The appropriation of Swahili in the former Belgian Congo, 1880–1938*. Cambridge: Cambridge University Press.
- Faggin, Giorgio. 1997. *Grammatica Friulana*. Campofornido, Italy: Ribis.
- Faltz, Leonard M. 1985. *Reflexivization: A study in universal syntax*. New York: Garland.
- Ferguson, Charles. 1959. The Arabic koine. *Language* 35:616–30.
- Ferlus, Michel. 1992. Essai de phonétique historique du khmer (du milieu du premier millénaire de notre ère à l’époque actuelle). *Mon-Khmer Studies* 21:57–89.
- Ferraz, Luis Ivens. 1984. Fanakalo: A pidgin caught in a crisis. In *York Papers in linguistics* 11, ed. Mark Sebba and Loreto Todd, 107–16. York: York Papers in Linguistics.
- Fischer, Olga. 1992. Syntax. In *The Cambridge history of the English language*, vol. 2, ed. Norman Blake, 207–408. Cambridge: Cambridge University Press.
- . 1997. Infinitive marking in Late Middle English: Transitivity and changes in the English system of case. In *Studies in Middle English linguistics*, ed. Jacek Fisiak, 109–34. Berlin: Mouton de Gruyter.
- Fischer, Olga, Ans van Kemenade, Willem Koopman, and Wim van der Wurff, eds. 2000. *The syntax of early English*. Cambridge: Cambridge University Press.

- Fischer, Wolfdietrich, and Otto Jastrow, eds. 1980. *Handbuch der arabischen Dialekte*. Wiesbaden: Otto Harrassowitz.
- Fisiak, Jacek, ed. 1995. *Linguistic change under contact conditions*. Berlin: Mouton de Gruyter.
- . 1997. *Studies in Middle English linguistics*. Berlin: Mouton de Gruyter.
- Foley, William A., and Robert Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Forster, P., R. Harding, A. Torroni, and H. J. Bandelt 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *American Journal of Human Genetics* 59:935–45.
- Fox, James J. 1982. The Rotinese chotbah as a linguistic performance. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 3, ed. Amran Halim, Lois Carrington, and Stephen A. Wurm, 311–18. Canberra: Pacific Linguistics.
- Franke, Herbert, and Denis Twitchett, eds. 1994. Introduction. In *The Cambridge history of China*, vol. 6, 1–42. Cambridge: Cambridge University Press.
- Friberg, Timothy. 1995. Konjo. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 563–71. Berlin: Mouton de Gruyter.
- Frye, Richard N. 1963. *The heritage of Persia*. Cleveland: World.
- Geertz, Clifford. 1968. Linguistic etiquette. In *Readings in the sociology of language*, ed. Joshua Fishman, 282–95. The Hague: Mouton.
- Gensler, Orin D. Forthcoming. *The Celtic-North African linguistic link: Substrata and typological argumentation*. Oxford: Oxford University Press.
- Gershevitch, Ilya, ed. 1985. *The Cambridge history of Iran*, vol. 2. Cambridge: Cambridge University Press.
- Gil, David. 1994. The Structure of Riau Indonesian. *Nordic Journal of Linguistics* 17:179–200.
- . 2001. Creoles, complexity, and Riau Indonesian. *Linguistic Typology* 2/3:325–71.
- . 2002. The prefixes di- and N- in Malay/Indonesian dialects. In *The history and typology of Western Austronesian voice systems*, ed. Fay Wouk and Malcolm Ross, 241–83. Canberra: Pacific Linguistics.
- . 2005. Isolating-Monocategorical-Associational Language. In *Categorization in Cognitive Science*, ed. H. Cohen and C. Lefebvre. Amsterdam: Elsevier.
- Gildea, Spike. 1993. The development of tense markers from demonstrative pronouns in Panare (Cariban). *Studies in Language* 17:53–73.
- Givón, Talmy. 1979. *Understanding Grammar*. New York: Academic Press.
- Good, Jeff. 2003a. Clause combining in Chechen. *Studies in Language* 27: 113–70.
- . 2003b. Tonal morphology in a creole: High-tone raising in Saramaccan serial verb constructions. In *Yearbook of morphology*, ed. Geert Booij and Jaap van Marle. Dordrecht: Kluwer.
- Gordon, E. V. 1927. *An introduction to Old Norse*. Oxford: Clarendon.
- Gorelova, Liliya M. 2002. *Manchu grammar*. Leiden: Brill.
- Görlach, Manfred. 1991. *Introduction to Early Modern English*. Cambridge: Cambridge University Press.
- Gould, Stephen Jay. 1996. *Full house: The spread of excellence from Plato to Darwin*. New York: Harmony.
- Goyette, Stephane. 2000. From Latin to Early Romance: A case of partial creolization? In *Language change and language contact in pidgins and Creoles*, ed. John H. McWhorter, 103–31. Amsterdam: John Benjamins.
- Grand'Henry, Jacques. 1976. *Les parlers arabes de la région du Mzāb*. Leiden: Brill.
- Grant, Anthony P. 1996. The evolution of functional categories in Grand Ronde Chinook Jargon: Ethnolinguistic and grammatical considerations. In *Changing meanings, changing functions*, ed. Philip Baker and Anand Syea, 225–42. London: University of Westminster Press.

- Greenberg, Joseph. 1966a. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- . 1966b. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Grammar*, ed. Joseph H. Greenberg, 73–113. Cambridge: Cambridge University Press.
- Grierson, George A. 1920. *Ishkashmi, Zebaki, and Yazghulami*. London: Royal Asiatic Society.
- Grijns, Cornelius D. 1991. *Jakarta Malay: A multidimensional approach to spatial variation*. Leiden: KITLV Press.
- Grotzfeld, Heinz. 1965. *Syrisch-Arabische Grammatik*. Wiesbaden: Otto Harrassowitz.
- Guisso, Richard W. L. 1979. The reigns of the empress Wu, Chung-tsung and Jui-tsung. In *The Cambridge history of China*, vol. 3, ed. Denis Twitchett John K. Fairbank, 290–332. Cambridge: Cambridge University Press.
- Gumperz, John J., and Stephen C. Levinson, eds. 1996. *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.
- Haberland, Hartmut. 1994. Danish. In *The Germanic languages*, ed. Ekkehard König and Johan van der Auwera, 313–48. London: Routledge.
- Habertzettl, Stefanie. 2005. *Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache*. Tübingen: Niemeyer.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59:781–819.
- Håkansson, Gisela, Manfred Pienemann, and S. Sayehli. 2002. Transfer and typological proximity in the context of L2 processing. *Second Language Research* 18:250–73.
- Ham, William H. 1999. Tone sandhi in Saramaccan: A case of substrate transfer? *Journal of Pidgin and Creole Languages* 14:45–91.
- Hamp, Eric. 1965. The Albanian dialect of Mëndres. *Die Sprache* 11:137–54.
- Harasowska, Marta. 1999. *Morphophonemic variability, productivity, and change: The case of Rusyn*. Berlin: Mouton de Gruyter.
- Harris, Alice, and Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Hashimoto, Mantaro. 1973. *The Hakka dialect: A linguistic study of its phonology, syntax and lexicon*. Cambridge: Cambridge University Press.
- . 1986. The Altaicization of Northern Chinese. In *Contributions to Sino-Tibetan Studies*, ed. John McCoy and Timothy Light, 76–97. Leiden: Brill.
- . 1987. Question and presupposition in Chinese. *Wang Li memorial volumes*, ed. Chinese Language Society of Hong Kong, 147–77. Hong Kong: Joint.
- Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20:271–87.
- . 1999. External possession in a European areal perspective. In *External possession*, ed. Doris L. Payne and Immanuel Barshi, 109–35. Amsterdam: John Benjamins.
- Haugen, Einar. 1976. *The Scandinavian languages: An introduction to their history*. London: Faber & Faber.
- . 1981. Language fragmentation in Scandinavia: Revolt of the minorities. In *Minority languages today*, ed. Einar Haugen, 100–119. Edinburgh: University Press.
- Hawkins, John A. 1985. *A comparative typology of English and German: Unifying the contrasts*. Austin: University of Texas Press.
- Heath, Jeffrey. 1975. Some functional relationships in grammar. *Language* 51:89–104.
- . 1978. *Linguistic diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- . 1981. A case of intensive lexical diffusion. *Language* 57:335–67.
- . 1997. Lost wax: Abrupt replacement of key morphemes in Australian agreement complexes. *Diachronica* 14:197–232.

- Hedevind, Bertil. 1967. *The dialect of Dentsdale in the West Riding of Yorkshire*. Uppsala: Appelbergs.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. From cognition to grammar: Evidence from African languages. In *Approaches to grammaticalization*, vol. 1, ed. Elizabeth Closs Traugott and Bernd Heine, 149–87. Amsterdam: John Benjamins.
- Henne, Henry. 1966. A sketch of Sathewkok Hakka grammatical structure. *Acta Linguistica Hafniensia* 10:69–108. Copenhagen: Nordisk Sprog-og Kulturforlag.
- Henry, Alison, and Denise Tangney. 1999. Functional categories and parameter setting in the second-language acquisition of Irish in early childhood. In *Language creation and language change*, ed. Michel DeGraff, 239–53. Cambridge, Mass.: MIT Press.
- Hermann, Eduard. 1895. “Gab es im Indogermanischen Nebensätze?” *Zeitschrift für vergleichende Sprachforschung* 33:481–535.
- Hetzron, Robert. 1987. The Semitic languages. In *The world’s major languages*, ed. Bernard Comrie, 654–63. New York: Oxford University Press.
- . ed. 1997. *The Semitic languages*. London: Routledge.
- Heusler, Andreas. 1950. *Altisländisches Elementarbuch*. Heidelberg: Carl Winter.
- Hill, Jane. 1993. Formalism, functionalism, and the discourse of evolution. In *The role of theory in language description*, ed. William Foley, 437–55. Berlin: Mouton de Gruyter.
- Hiltunen, Risto. 1983. *The decline of prefixes and the beginnings of the English phrasal verb*. Turku: Turun Yliopisto.
- Ho, Mian-Lian, and John T. Platt. 1993. *Dynamics of a contact continuum*. Oxford: Clarendon.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Berlin: Mouton de Gruyter.
- Hodge, Carlton T. 1970. The linguistic cycle. *Language Sciences* 13:1–7.
- Holes, Clive. 1984. *Colloquial Arabic of the Gulf and Saudi Arabia*. London: Routledge.
- . 1990. *Gulf Arabic*. London: Routledge.
- Holm, John. 1986. Substrate diffusion. In *Substrata versus universals in creole genesis*, ed. Pieter Muysken and Norval J. Smith, 259–78. Amsterdam: John Benjamins.
- . 1987. Creole influence on Popular Brazilian Portuguese. In *Pidgin and creole languages*, ed. Glenn G. Gilbert, 406–29. Honolulu: University of Hawaii Press.
- . 1989. *Pidgins and creoles*, vol. 2. Cambridge: Cambridge University Press.
- . 2003. *Languages in contact: The partial restructuring of vernaculars*. Cambridge: Cambridge University Press.
- Holmes, Philip, and Ian Hinchliffe. 1997. *Swedish: An essential grammar*. London: Routledge.
- Hope, Jonathan. 1994. The use of *thou* and *you* in Early Modern spoken English: Evidence from depositions in the Durham ecclesiastical court records. In *Studies in Early Modern English*, ed. Dieter Kastovsky, 141–51. Berlin: Mouton de Gruyter.
- Hopper, Paul J. 1975. *The syntax of the simple sentence in Proto-Germanic*. The Hague: Mouton.
- Hopper, Paul J., and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Horne, Elinor C. 1961. *Beginning Javanese*. New Haven, Conn.: Yale University Press.
- Huang, Zaijun. 1964. Cong jiawen jinwen liangci-de yingyong kaocha hanyu liangci-de quyan yu fazhan [A study of the origin and development of classifiers in Chinese, beginning from their use in the oracle bone inscriptions and metal inscriptions]. *Zhongguo Yuwen* 133(6): 432–41. Beijing: Beijing Renmin Jiaoyu Chubanshe.
- Huffman, Franklin E. 1970. *Modern spoken Cambodian*. New Haven, Conn.: Yale University Press.
- Hull, Geoffrey. 1998. The basic lexical affinities of Timor’s Austronesian languages: A preliminary investigation. *Studies in the Languages and Cultures of East Timor* 1:97–174.

- . 2001. A morphological overview of the Timoric Sprachbund. *Studies in Languages and Cultures of East Timor* 4:98–205.
- Humbach, Helmut. 1989. Choresmian. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 193–203. Wiesbaden: Ludwig Reichert.
- Hyman, Larry. Forthcoming. How to become a Kwa verb. *Journal of West African languages*.
- Ingham, Bruce. 1992. *North East Arabian dialects*. London: Kegan Paul.
- . 1994. *Najdi Arabic*. Amsterdam: John Benjamins.
- Jacobsson, Bengt. 1951. *Inversion in English: With special reference to the Early Modern English period*. Uppsala: Almqvist & Wiksell.
- Jahr, Ernst Håkon. 2001. Historical sociolinguistics: The role of Low German language contact in the Scandinavian typological shift of the late Middle Ages. *Lingua Posnaniensis* 43:95–104.
- Jastrow, Otto. 1997. The Neo-Aramaic Languages. In *The Semitic languages*, ed. Robert Hetzron, 334–77. London: Routledge.
- Johnstone, T. M. 1967. *Eastern Arabian dialect studies*. London: Oxford University Press.
- Jonker, J. C. G. 1915. *Rottineesche spraakkunst*. Leiden: Brill.
- Kapanga, André Mwamba. 1993. Shaba Swahili and the processes of linguistic contact. In *Atlantic meets Pacific: A global view of pidginization and creolization*, ed. Francis Byrne and John Holm, 441–58. Amsterdam: John Benjamins.
- Karamkhudoev, H. 1973. *Bartangskiy jazyk*. Dushanbe, Tajikistan: Donish.
- Kastovsky, Dieter. 1992. Semantics and vocabulary. *The Cambridge history of the English language*, vol. 1, ed. Richard M. Hogg, 290–408. Cambridge: Cambridge University Press.
- Kaye, Alan. 1976. *Chadian and Sudanese Arabic in the light of comparative Arabic dialectology*. The Hague: Paris.
- . 1986. The verb “see” in Arabic dialects. In *The Fergusonian impact*, vol. 1, ed. Joshua Fishman, 211–21. Berlin: Mouton De Gruyter.
- . 1987. Arabic. In *The world’s major languages*, ed. Bernard Comrie, 664–85. New York: Oxford University Press.
- Kaye, Alan, and Judith Rosenhouse. 1997. Arabic dialects and Maltese. In *The Semitic languages*, ed. Robert Hetzron, 263–311. London: Routledge.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.
- Keeler, Ward. 1984. *Javanese: A cultural approach*. Athens: Ohio University Center for International Studies.
- Keesing, Roger M. 1985. *Kwaio grammar*. Canberra: Pacific Linguistics.
- . 1988. *Melanesian Pidgin and the Oceanic substrate*. Palo Alto, Calif.: Stanford University Press.
- Keller, Rudi. 1994. *On language change: The invisible hand in language*. London: Routledge.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.
- Kent, Roland G. 1953. *Old Persian*. New Haven, Conn.: American Oriental Society.
- Khan, Emir Djeladet Bedir, and Roger Lescot. 1970. *Grammaire kurde (dialecte kurmandji)*. Paris: Librairie D’Amérique et d’Orient.
- Khanlari, P. N. 1979. *A history of the Persian language*. Delhi: Idarah-i Adabiyat-i Delli.
- Khromov, A. L. 1972. *Yagnobskii Jazyk*. Moscow: Nauka.
- Kieffer, Charles M. 2003. *Grammaire de l’ormuṛī de Baraki-Barak*. Wiesbaden: Ludwig Reichert.
- Kilpiö, M. 1989. *Passive constructions in Old English translations from Latin: With special reference to the OE Bede and the “Pastoral Care.”* Helsinki: Société Néophilologique.
- Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In *Clause structure and*



- language change*, ed. Adrian Battye and Ian Roberts, 140–69. New York: Oxford University Press.
- Kiparsky, Valentin. 1969. *Gibt es ein Finnougrisches Substrat im Slavischen?* Helsinki: Suomalainen Tiedekatemia.
- Klamer, Marian. 1998. *A grammar of Kambara*. Berlin: Mouton de Gruyter.
- Klemola, K. Juhani. 2000. The origins of the Northern Subject Rule: A case of early contact? In *The Celtic Englishes II*, ed. Hildegard L. C. Tristram, 329–46. Heidelberg: C. Winter.
- Kloetze, G. G. 1950. *Herkomst en groei van het Afrikaans*. Leiden: Universitaire Pers.
- König, Ekkehard, and Johan van der Auwera, eds. 1994. *The Germanic languages*. London: Routledge.
- and Martin Haspelmath. 1997. Les constructions à possesseur externe dans les langues d'Europe. In *Actance et valence dans les langues de l'Europe*, ed. Jack Feuillet, 525–606. Berlin: Mouton de Gruyter.
- Kotsinas, Ulla-Britt. 2001. Pidginization, creolization and creoloids in Stockholm, Sweden. In *Creolization and contact*, ed. Norval Smith and Tonjes Veenstra, 125–55. Amsterdam: John Benjamins.
- Kouwenberg, Silvia. 1994. *A grammar of Berbice Dutch Creole*. Berlin: Mouton de Gruyter.
- Kramer, Marvin. 2002. Substrate transfer in Saramaccan creole. PhD diss., University of California, Berkeley.
- Kress, Bruno. 1982. *Isländische Grammatik*. Munich: Max Hüder.
- Kroch, Anthony, and Ann Taylor. 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In *Parameters of morphosyntactic change*, ed. Ans Van Kemenade and Nigel Vincent, 297–325. Cambridge: Cambridge University Press.
- Kubler, Cornelius C. 1985. *The development of Mandarin in Taiwan: A case study of language contact*. Taipei: Student Book.
- Kumudawati, Ratri, and Theresa Warjanti Setjadiningrat. 2001. Distinguishing features of Ambon and Kupang Malay. *Studies in Languages and Cultures of East Timor* 4:61–98.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: Landelijke Onderzoeksschool Taalwetenschap [Netherlands Graduate School of Linguistics].
- LaCroix, Pierre-François. 1967. *Quelques aspects de la désintégration d'un système classificatoire (peul du sud de l'Adamawa). La classification nominale dans les langues négro-africaines (Colloques internationaux du Centre National de la Recherche Scientifique)*. Paris: Éditions du Centre National de la Recherche Scientifique.
- Lamarre, Christine. 2001. Verb-complement constructions in Chinese dialects: Types and markers. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 85–120. New York: Oxford University Press.
- LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In *Areal diffusion and genetic inheritance*, ed. Alexandra Y. Aikhenvald and R. M. W. Dixon, 225–54. New York: Oxford University Press.
- Lass, Roger. 1987. *The shape of English*. London: J. M. Dent & Sons.
- . 1990. How to do things with junk: Exaptation in language evolution. *Journal of Linguistics* 26:79–102.
- . 1992. Phonology and morphology. In *The Cambridge history of the English language*, vol. 2, ed. Norman Blake, 23–155. Cambridge: Cambridge University Press.
- Lau, Chun-fat. 1999. "Gender" in the Hakka dialect: Suffixes with gender in more than 40 nouns. *Journal of Chinese Linguistics* 27:124–31.
- Lazard, Gilbert. 1975. The rise of the new Persian language. In *The Cambridge history of Iran*, vol. 4, ed. Richard N. Frye, 595–632. Cambridge: Cambridge University Press.

- . 1989. Le persan. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 263–93. Wiesbaden: Ludwig Reichert.
- . 1992. *A grammar of contemporary Persian*. Costa Mesa, Calif.: Mazda.
- LeCoq, Pierre. 1989. Les dialectes du centre de l'Iran. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 313–26. Wiesbaden: Ludwig Reichert.
- Lee-Smith, Mei W. 1996a. The Tangwang language. In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*, vol. 2.2, ed. Stephen A. Wurm, Peter Mühlhäusler, and Darrell T. Tyron, 875–82. Berlin: Mouton de Gruyter.
- . 1996b. The Hezhou language. In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*, vol. 2.2, ed. Stephen A. Wurm, Peter Mühlhäusler, and Darrell T. Tyron, 865–73. Berlin: Mouton de Gruyter.
- Lee-Smith, Mei W., and Stephen A. Wurm. 1996. The Wutun language. In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*, vol. 2.2, ed. Stephen A. Wurm, Peter Mühlhäusler, and Darrell T. Tyron, 883–97. Berlin: Mouton de Gruyter.
- Lefebvre, Claire. 1998. *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.
- . 2004. *Issues in the study of pidgin and creole languages*. Amsterdam: John Benjamins.
- Lefebvre, Claire, and Anne-Marie Brousseau. 2001. *A grammar of Fongbe*. Berlin: Mouton de Gruyter.
- Lewis, E. D. and Charles E. Grimes. 1995. Sika. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 601–9. Berlin: Mouton de Gruyter.
- Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley and Los Angeles: University of California Press.
- . 1982. The gulf between spoken and written language: A case study in Chinese. In *Spoken and written language: Exploring orality and literacy*, ed. Deborah Tannen, 77–88. Norwood, N.J.: Ablex.
- . 1987. Chinese. In *The world's major languages*, ed. Bernard Comrie, 811–33. New York: Oxford University Press.
- Li, Ying-Che. 1986. Historical significance of certain distinct grammatical features in Taiwanese. In *Contributions to Sino-Tibetan Studies*, ed. John McCoy and Timothy Light, 393–414. Leiden: Brill.
- . 2001. Aspects of historical-comparative syntax: Functions of prepositions in Taiwanese and Mandarin. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 340–68. Oxford: Oxford University Press.
- Lien, C. 2001. Competing morphological changes in Taiwanese Southern Min. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 309–39. Oxford: Oxford University Press.
- Lightfoot, David. 1997. Shifting triggers and diachronic reanalyses. In *Parameters of morphosyntactic change*, ed. Ans Van Kemenade and Nigel Vincent, 253–72. Cambridge: Cambridge University Press.
- . 1999. *The development of language*. Oxford: Blackwell.
- Lockwood, W. B. 1955. *An introduction to Modern Faroese*. Copenhagen: Ejnar Munksgaard.
- . 1995. *Lehrbuch der modernen jiddischen Sprache*. Hamburg: Helmut Buske.
- Lorimer, D. L. R. 1958. *The Wakhi language*, vol. 1. London: School of Oriental and African Studies.
- Lück, M., and L. Henderson. 1993. *Gambian Mandinka*. Banjul, Gambia: WEC International.
- Lucy, John A. 1992. *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.



- Luo, Zhaojin. 1984. *Keyu yufa* [Hakka grammar]. Taipei: Student Book Company.
- Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Macdonald, R. Ross. 1976. *Indonesian reference grammar*. Washington, D.C.: Georgetown University Press.
- MacKenzie, D. N. 1987. Pashto. In *The world's major languages*, ed. Bernard Comrie, 547–65. New York: Oxford University Press.
- Marcantonio, Angela. 2002. *The Uralic language family: Facts, myths, and statistics*. Oxford: Blackwell.
- Marchand, Hans. 1969. *The categories and types of present-day English word formation*. Munich: Beck.
- Martens, Michael. 1995. Uma. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 537–47. Berlin: Mouton de Gruyter.
- Mathes, B. F. 1858. *Makassaarsche Spraakkunst*. Amsterdam: C. A. Spin & Zoon.
- Matisoff, James A. 1973. *The Grammar of Lahu*. Berkeley: University of California Publications.
- . 1991. Areal and universal dimensions of grammaticization in Lahu. In *Approaches to Grammaticalization*, vol. 2, ed. Elizabeth C. Traugott and Bernd Heine, 383–453. Amsterdam: John Benjamins.
- . 2001. Genetic versus contact relationship: Prosodic diffusibility in South-East Asian languages. In *Areal diffusion and genetic inheritance*, ed. Alexandra Y. Aikhenvald and R. M. Dixon, 291–327. New York: Oxford University Press.
- Matthews, Stephen, and Virginia Yip. 1994. *Cantonese: A comprehensive grammar*. London: Routledge.
- . 2001. Aspects of contemporary Chinese grammar: The structure and stratification of relative clauses. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 266–82. Oxford: Oxford University Press.
- Matthews, William K. 1956. *The Livonian element in modern Latvian*. Festschrift für Max Vasmer zur 70. Geburtstag, ed. Margarete Woltner and Herbert Bräuer, 307–18. Berlin: Veröffentlichungen der Abteilung für slawische Sprachen und Literaturen des Osteuropa-Instituts an der Freien Universitäts Berlin.
- McCoy, John, and Timothy Light, eds. 1986. *Contributions to Sino-Tibetan Studies*. Leiden: Brill.
- McWhorter, John H. 1994. From focus marker to copula in Swahili. *Special session on historical issues in African Languages: Proceedings of the 20th annual meeting of the Berkeley Linguistics Society*, ed. Kevin E. Moore, David A. Peterson, and Comfort Wentum, 57–66. Berkeley, Calif.: Berkeley Linguistics Society.
- . 1998. The word on the street: Fact and fable about American English. New York: Plenum.
- , ed. 2000a. *Language change and language contact in pidgins and creoles*. Amsterdam: John Benjamins.
- . 2000b. *The missing Spanish creoles: Recovering the birth of plantation contact languages*. Berkeley and Los Angeles: University of California Press.
- . 2000c. Strange bedfellows: Recovering the origin of Black English: Review article of *The English History of African American English*, ed. Shana Poplack. *Diachronica* 17:389–432.
- . 2001. The rest of the story: Restoring pidginization to creole genesis theory. *Journal of Pidgin and Creole Languages* 17:1–48.
- . 2002. What happened to English? *Diachronica* 19:217–72.

- . 2004. Saramaccan and Haitian as young grammars: The pitfalls of syntactocentrism in creole genesis research. *Journal of Pidgin and Creole Languages* 19: 77–138.
- . 2005. *Defining creole*. New York: Oxford University Press.
- McWhorter, John H., and Mikael Parkvall. 2002. Pas tout à fait du français: Une étude créole. *Études Créoles* 25:179–231.
- Meillet, A. 1912. Les nouvelles langues indo-européennes trouvées en Asie Centrale. *Revue du mois* 14:135–52.
- Meier, Hans Heinrich. 1953. *Der indefinite Agens im Mittelhochdeutschen (1050–1350): Die Wörter und Wendungen für “man.”* Bern: A. Francke.
- Melchers, Gunnel. 1992. “Du’s no heard da last o’ dis” — on the use of *be* as a perfective auxiliary in Shetland dialect. In *History of Englishes*, ed. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen, and Irma Taavitsainen, 602–10. Berlin: Mouton de Gruyter.
- Meroney, Howard M. 1945. The early history of “down” as an adverb. *Journal of English and Germanic Philology* 44:378–86.
- Mitchell, Bruce. 1985. *Old English syntax*, vol. 1. Oxford: Clarendon.
- and Fred C. Robinson. 1986. *A guide to Old English*. Oxford: Blackwell.
- Mitchell, T. F. 1956. *An introduction to Egyptian colloquial Arabic*. London: Oxford University Press.
- Mittendorf, Ingo, and Erich Poppe. 2000. Celtic contacts of the English progressive? In *The Celtic Englishes II*, ed. Hildegard L. C. Tristram, 117–45. Heidelberg: C. Winter.
- Mittwoch, Anita. 1990. On the distribution of infinitive complements in English. *Journal of Linguistics* 26: 103–31.
- Molchanova, E. K. 1997. Sivendi. In *Osnovy iranskovo jazykoznanija: Novoiranskije jazyki: Severo-zapadnaja grupa*, vol. 2, ed. B. A. Efimov, 330–417. Moscow: Vostočnaja literatura R. A. N.
- Moravcsik, Edith. 1969. Determination. Working Papers on Language Universals I, 64–130. Palo Alto, Calif.: Stanford University Press.
- Morris, Simon Conway. 2003. *Life’s solution*. Cambridge: Cambridge University Press.
- Moussay, Gerard. 1981. *La langue minangkabau*. Paris: Archipel.
- Müller-Gotama, Franz. 2001. *Sundanese*. Munich: Lincom Europa.
- Mustanoja, Tauno F. 1960. *A Middle English syntax*, vol. 1. Helsinki: Société Néophilologique.
- Muysken, Pieter. 1997. Media Lengua. In *Contact languages: A wider perspective*, ed. Sarah G. Thomason, 365–426. Amsterdam: John Benjamins.
- Muysken, Pieter, and Norval J. Smith, eds. 1986. *Substrata versus universals in creole genesis*. Amsterdam: John Benjamins.
- Nababan, P. W. J. 1981. *A grammar of Toba Batak*. Canberra: Australian National University.
- Nagaraja, K. S. 1985. *Khasi: A descriptive analysis*. Deccan College Post-Graduate and Research Institute: Pune 6.
- Naro, Anthony J., and Maria Mara Pereira Scherre. 2000. Variable concord in Portuguese: The situation in Brazil and Portugal. In *Language change and language contact in pidgins and Creoles*, ed. John H. McWhorter, 235–55. Amsterdam: John Benjamins.
- Nevalainen, Terttu. 1997. Recycling inversion: The case of initial adverbs and negators in Early Modern English. A Festschrift for Roger Lass on his sixtieth birthday *Studia Anglica Posnaniensa* 32:203–14.
- Newmeyer, Frederick J. 1986. *The politics of linguistics*. Chicago: University of Chicago Press.
- Norman, Jerry. 1973. Tonal development in Min. *Journal of Chinese Linguistics* 1:222–38.
- . 1988. *Chinese*. Cambridge: Cambridge University Press.

- . 2003. The Chinese dialects: Phonology. In *The Sino-Tibetan languages*, ed. Graham Thurgood and Randy J. LaPolla, 72–83. London: Routledge.
- Nothofer, Bernd. 1975. *The reconstruction of Proto-Malayo-Javanic*. The Hague: Martinus Nijhoff.
- Nurse, Derek. 1997. Prior pidginization and creolization in Swahili? In *Contact languages: A wider perspective*, ed. Sarah G. Thomason, 271–94. Amsterdam: John Benjamins.
- Ogawa, Tamaki. 1981. Suzhou fangyan de zhishi daici [Demonstratives in the Suzhou dialect]. *Fangyan* 4:287–88.
- O'Grady, William, Michael Dobrovolsky, and Mark Aronoff. 1997. *Contemporary linguistics: An introduction*. 3rd ed. New York: St. Martin's.
- Omar, Asmah Haji. 1981. *The Iban language of Sarawak*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- O'Neil, Wayne. 1978. The evolution of the Germanic inflectional systems: A study in the causes of language change. *Orbis* 27:248–86.
- Oppenheimer, Stephen. 2003. *The real Eve: Modern man's journey out of Africa*. New York: Carroll & Graf.
- Owens, Jonathan. 1984. *A short reference grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrassowitz.
- . 1993. *A grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz.
- . 1997. Arabic-based pidgins and creoles. In *Contact languages: A wider perspective*, ed. Sarah J. Thomason, 125–72. Amsterdam: John Benjamins.
- Pagel, Karl. 1983. *Die Hanse*. Braunschweig: Westermann.
- Pakhalina, T. N. 1991. Semnani. In *Osnovy iranskovo jazykoznanija: Novoiranskije jazyki: Severo-zapadnaja gruppa*, vol. 1, ed. V. S. Rastorgueva, 176–204.
- Paul, Hermann. 1880. *Principien der Sprachgeschichte*. Halle: Niemeyer.
- Payne, John. 1989. Pāmir languages. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 417–44. Wiesbaden: Ludwig Reichert.
- Payne, Thomas E. 1997. *Describing morphosyntax*. Cambridge: Cambridge University Press.
- Peitsara, Kirsti. 1997. The development of reflexive strategies in English. In *Grammaticalization at work*, ed. Matti Rissanen, Merja Kytö, and Kirsi Heikkinen, 277–370. Berlin: Mouton de Gruyter.
- Perkins, Revere D. 1992. *Deixis, grammar, and culture*. Amsterdam: John Benjamins.
- Peters, Robert. 1987. Das Mittelniederdeutsche als Sprache der Hanse. In *Sprachkontakt in der Hanse: Aspekte des Sprachausgleichs im Ostsee-und Nordseeraum*, ed. Sture Ureland, 65–88. Tübingen: Niemeyer.
- Pienemann, Manfred, Bruno Di Biase, Satomi Kawaguchi, and Gisela Häkansson. Forthcoming. Processing constraints on L1 transfer. In *Handbook of bilingualism: Psycholinguistic approaches*, ed. J. F. Kroll and A. M. B. DeGroot. New York: Oxford University Press.
- Pintzuk, Susan. 1991. Phrase structures in competition: Variation and change in Old English word order. PhD diss., University of Pennsylvania.
- Pireiko, L.A. 1997. Gurani. In *Osnovy iranskovo jazykoznanija: Novoiranskije jazyki: Severo-zapadnaja gruppa*, vol. 2, ed. B. A. Efimov, 144–93. Moscow: Vostočnaja literatura R. A. N.
- Plank, Frans, and Edith Moravcsik. 1996. The Maltese article: Language-particulars and universals. *Rivista di Linguistica* 8:183–212.
- Platt, John T. 1975. The Singapore English speech continuum and its basilect “Singlish” as a “creoloid.” *Anthropological Linguistics* 17:363–74.
- Platzack, Christer. 1986. COMP, INFL, and Germanic word order. In *Topics in Scandinavian syntax*, ed. Lars Hellan and Kirsti Koch Christensen, 185–234. Dordrecht: Reidel.

- Platzack, Christer, and Anders Holmberg. 1989. The role of AGR and finiteness in Germanic VO languages. *Working Papers in Scandinavian Syntax* 43:51–76.
- Poedjosoedarmo, Gloria R. 2002. Changes in word order and noun phrase marking from Old to modern Javanese: Implications for understanding developments in western Austronesian “focus” systems. In *The history and typology of Western Austronesian voice systems*, ed. Fay Wouk and Malcolm Ross, 311–30. Canberra: Pacific Linguistics.
- Polome, Edgar C. 1968. Lubumbashi Swahili. *Journal of African Languages* 7: 14–25.
- Ponelis, Fritz. 1993. *The development of Afrikaans*. Frankfurt: Peter Lang.
- Poussa, Patricia. 1982. The evolution of early standard English: The creolization hypothesis. *Studia Anglica Posnaniensia* 14:69–85.
- . 1990. A contact-universals explanation for periphrastic *do*, with special consideration of OE-Celtic contact. In *Papers from the 5th International Conference on English Historical Linguistics*, ed. Sylvia Adamson, Vivien A. Law, Nigel Vincent, and Susan Wright, 407–34. Amsterdam: John Benjamins.
- Prentice, D. J. 1987. Malay (Indonesian and Malaysian). In *The world's major languages*, ed. Bernard Comrie, 913–35. New York: Oxford University Press.
- Prentice, D. J., and A. Hakim Usman. 1978. Kerinci sound changes and phonotactics. In *Second International Conference on Austronesian Linguistics: Proceedings*, fascicle 1, ed. Stephen A. Wurm and Lois Carrington, 121–63. Canberra: Pacific Linguistics.
- Quick, Phil. 2002. Creaky voice as a phonetic manifestation of the glottal stop in Pendau. In *Proceedings of AFLA 8: The eighth meeting of the Austronesian Formal Linguistics Association*, ed. Andrea Rackowski and Norvin Richards, 293–307. Cambridge, Mass.: MIT Working Papers in Linguistics.
- Rabel-Heyman, Lili. 1977. Gender in Khasi nouns. *Mon-Khmer Studies* 6:247–72.
- Ramat, Paolo. 1998. The Germanic languages. In *The Indo-European languages*, ed. Anna Giacalone Ramat and Paolo Ramat, 380–414. London: Routledge.
- Rastorgueva, V. S. 1991a. Sangesari. In *Osnovy iranskovo jazykoznaniya: Novoiranskije jazyki: Severo-zapadnaja gruppa*, vol. 1, ed. V. S. Rastorgueva, 206–46. Moscow: Vostoc&naja literatura R. A. N.
- . ed. 1991b. *Osnovy iranskovo jazykoznaniya: Novoiranskije jazyki: Severo-zapadnaja gruppa*, vol. 1. Moscow: Vostoc&naja literatura R. A. N.
- Rastorgueva, V. S., A. A. Kerimova, A. K. Mamedzade, L. A. Pireiko, and D. N. Edelman. 1971. *Giljanskiy jazyk*. Moscow: Nauka.
- Rastorgueva, V. S., and B. B. Moshkalo. 1997. Dialekty tsentral’no irana. In *Osnovy iranskovo jazykoznaniya: Novoiranskije jazyki: Severo-zapadnaja gruppa*, vol. 2, ed. B. A. Efimov, 195–326. Moscow: Vostoc&naja literatura R. A. N.
- Renfrew, Colin. 1987. *Archaeology and language*. Cambridge: Cambridge University Press.
- Rhodes, Richard. 1990. Obviation, inversion, and topic rank in Ojibwa. In *Proceedings of the Berkeley Linguistics Society* 16, ed. David J. Costa, 101–15. Berkeley: University of California, Berkeley.
- Rissanen, Matti. 1997. Whatever happened to the Middle English indefinite pronouns? In *Studies in Middle English linguistics*, ed. Jacek Fisiak, 513–29. Berlin: Mouton de Gruyter.
- . 1999. *Syntax*. In *The Cambridge history of the English language*, vol. 1, ed. Roger Lass, 187–331. Cambridge: Cambridge University Press.
- Roberge, Paul. 2003. Review of degrees of restructuring in Creole languages, ed. Ingrid Neumann-Holzschuh and Edgar Schneider. *Journal of Pidgin and Creole Languages* 18:339–48.
- Roberts, Ian. 1993. *Verbs and diachronic syntax: A comparative history of English and French*. Dordrecht: Kluwer.

- . 1999. Verb movement and markedness. In *Language creation and language change*, ed. Michel DeGraff, 287–327. Cambridge, Mass.: MIT Press.
- Roberts, Sarah Julianne. 2000. Nativization and the genesis of Hawaiian Creole. In *Language change and language contact in pidgins and Creoles*, ed. John H. McWhorter, 257–300. Amsterdam: John Benjamins.
- Robinson, Ian. 1975. *The new grammarians' funeral: A critique of Noam Chomsky's linguistics*. Cambridge: Cambridge University Press.
- Rohrbacher, Bernhard. 1999. *Morphology-driven syntax: A theory of V to I raising and pro-drop*. Amsterdam: John Benjamins.
- Root, M. Cool. 1997. *Cultural pluralisms on the Persepolis Fortification tablets*. Topoi. Suppl. 1: Recherches récentes sur l'Empire achéménide, 229–52.
- Rosenhouse, Judith. 1984. *The Bedouin Arabic dialects*. Wiesbaden: Otto Harrassowitz.
- Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In *The history and typology of Western Austronesian voice systems*, ed. Fay Wouk and Malcolm Ross, 17–78. Canberra: Pacific Linguistics.
- Rountree, S. Catharine. 1972. Saramaccan tone in relation to intonation and grammar. *Lingua* 29:308–25.
- Rydén, Mats, and Sverker Brorström. 1987. *The “be/have” variation with intransitives in English: With special reference to the Late Modern period*. Stockholm: Almqvist & Wiksell.
- Sagart, Laurent. 2001. Vestiges of Archaic Chinese derivational affixes in Modern Chinese dialects. In *Sinitic grammar: Synchronic and diachronic perspectives*, ed. Hilary Chappell, 123–42. Oxford: Oxford University Press.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace Jovanovich.
- Sawyer, P. H. 1971. *The age of the Vikings*. London: Edward Arnold.
- Schachter, Paul. 1987. Tagalog. In *The world's major languages*, ed. Bernard Comrie, 936–58. New York: Oxford University Press.
- Schaller, Helmut W. 1975. *Die Balkansprachen: Ein Einführung in die Balkanphilologie*. Heidelberg: Carl Winter.
- Schmitt, Rüdiger. 1989a. Altpersisch. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 56–85. Wiesbaden: Ludwig Reichert.
- , ed. 1989b. *Compendium linguarum Iranicarum*. Wiesbaden: Ludwig Reichert.
- Schuchardt, Hugo. 1914. *Die Sprache der Saramakkaneger in Suriname*. Amsterdam: Johnannes Müller.
- Schulz, Eckehard, Günther Krah, and Wolfgang Reuschel. 2000. *Standard Arabic: An elementary-intermediate course*. Cambridge: Cambridge University Press.
- Schulze, Wolfgang. 2000. *Northern Talysh*. Munich: Lincom Europa.
- Schwegler, Armin. 1998. El palenquero. In *América negra: Panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, ed. Matthias Perl and Armin Schwegler, 219–91. Frankfurt: Vervuert.
- Selcan, Zülfi. 1998. *Grammatik der Zaza-Sprache*. Berlin: Wissenschaft und Technik.
- Siegel, Jeff. 1987. *Language contact in a plantation environment*. Cambridge: Cambridge University Press.
- . 2000a. Substrate influence in Hawai'i Creole English. *Language in Society* 29:197–236.
- , ed. 2000b. *Processes of language contact: Studies from Australia and the South Pacific*. Québec: AGMV Marquis.
- Siegel, Jeff, Barbara Sandeman, and Chris Corne. 2000. Predicting substrate influence: Tense-modality-aspect marking in Tayo. In *Processes of language contact: Studies from Australia and the South Pacific*, ed. Jeff Siegel, 75–97. Québec: AGMV Marquis.

- Simenone-Senelle, Marie-Claude. 1997. The modern South Arabian languages. In *The Semitic languages*, ed. Robert Hetzron, 378–423. London: Routledge.
- Simmons, Richard VanNess. 1992. The Hangzhou dialect. PhD diss., University of Washington.
- Sims-Williams, Nicholas. 1989a. Bactrian. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 230–35. Wiesbaden: Ludwig Reichert.
- . 1989b. Sogdian. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 173–92. Wiesbaden: Ludwig Reichert.
- . 1998. The Iranian languages. In *The Indo-European languages*, ed. Anna Giacalone Ramat and Paolo Ramat, 125–53. London: Routledge.
- Sinor, Denis. 1998. Central Asia (“Turkmenistan”). In *Encyclopedia Britannica* (Macropedia). Chicago: Encyclopedia Britannica.
- Sirk, U. H. 1979. La langue bugis. Trans. Nicole Pacory. Paris: Archipel.
- Skjaervø, Prods O. 1989a. Modern East Iranian languages. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 371–83. Wiesbaden: Ludwig Reichert.
- . 1989b. Yidgha and Munji. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 371–83. Wiesbaden: Ludwig Reichert.
- Smith, Henry. 1992. “Dative sickness” and abstractness. In *On Germanic linguistics: Issues and methods*, ed. Irmengard Rauch, Gerald F. Carr, and Robert L. Kyes, 283–97. Berlin: Mouton de Gruyter.
- Smith-Hefner, Nancy J. 1988. Cara Tengger: Notes on a non-standard dialect of Javanese. In *Studies in Austronesian linguistics*, ed. Richard McGinn, 203–33. Athens: Ohio University Center for International Studies.
- Sneddon, James Neil. 1996. *Indonesian: A comprehensive grammar*. London: Routledge.
- Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76:859–90.
- Spears, Arthur K., and Donald Winford, eds. 1997. The structure and status of pidgins and creoles. Amsterdam: John Benjamins.
- Spencer, Andrew. 1991. *Morphological theory*. Oxford: Blackwell.
- Srivastava, V. C. 1996. *The proto-historic Afghanistan: A source book*. Varanasi: Tara Book Agency.
- Steinhauer, Hein. 1993. Notes on verbs in Dawanese (Timor). In *Topics in descriptive Austronesian linguistics*, ed. Ger P. Reesink, 130–58. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië.
- Stevens, Alan. 1968. *Madurese phonology and morphology*. New Haven, Conn.: American Oriental Society.
- Strandskogen, Åse-Berit, and Rolf Strandskogen. 1986. *Norwegian: An essential grammar*. London: Routledge.
- Strang, Barbara M. H. 1970. *A history of English*. London: Methuen.
- Streitberg, Wilhelm. 1906. *Gotisches Elementarbuch*. Heidelberg: Carl Winter.
- Suharno, Ignatius. 1982. *A descriptive study of Javanese*. Canberra: Australian National University.
- Sundermann, Werner. 1989a. Mittelpersisch. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 138–64. Wiesbaden: Ludwig Reichert.
- . 1989b. Westmittliranische Sprachen. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 106–13. Wiesbaden: Ludwig Reichert.
- Sung, Margaret M. Y. 1973. A study of literary and colloquial Amoy Chinese. *Journal of Chinese Linguistics* 1:414–36.
- Sutcliffe, Edmund F. 1936. *A grammar of the Maltese language*. London: Oxford University Press.



- Szemerényi, O. 1980. Language decay — the result of imperial aggrandizement? In *Recherches de Linguistique: Hommages à Maurice leRoy*, ed., 206–14. Brussels: Editions de l'Université.
- Tadmor, Uri. 2003. The pronominal system of Classical Malay. Paper presented at the Seventh International Symposium on Malay/Indonesian linguistics, Nijmegen.
- Tchekhoff, C. 1979. From ergative to accusative in Tongan: An example of synchronic dynamics. In *Ergativity: Towards a theory of grammatical relations*, ed. Frans Plank, 407–18. London: Academic.
- Teselkin, A. S. 1972. *Old Javanese (Kawi)*. Trans. John M. Echols. Ithaca, N.Y.: Modern Indonesia Project, Cornell University.
- Thomas, E. J. 1991. *Danish conversation-grammar*. Heidelberg: Julius Groos.
- Thomason, Sarah J., ed. 1995. *Contact languages: a wider perspective*. Amsterdam: John Benjamins.
- . 1997a. A typology of contact languages. In *The structure and status of pidgins and creoles*, ed. Arthur K. Spears and Donald Winford, 71–88. Amsterdam: John Benjamins.
- . 1997b. *Contact languages: A wider perspective*. Amsterdam: John Benjamins.
- Thomason, Sarah J., and Alaa Elgibali. 1986. Before the Lingua Franca: Pidginized Arabic in the eleventh century A.D. *Lingua* 68:317–49.
- Thomason, Sarah J., and Terence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley and Los Angeles: University of California Press.
- Thongkum, Theraphan L. 1981. The distribution of the sounds of Bruu. *Mon-Khmer Studies* 8:221–93.
- Thordarson, Fridrik. 1989. Ossetic. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 456–79. Wiesbaden: Ludwig Reichert.
- Thurgood, Graham. 1999. *From ancient Cham to modern dialects: 2000 years of language contact and change*. Honolulu: University of Hawaii Press.
- Thurgood, Graham, and Randy J. LaPolla, eds. 2003. *The Sino-Tibetan languages*. London: Routledge.
- Thurston, William R. 1987. *Processes of change in the languages of northwestern New Britain*. Canberra: Australian National University.
- Tiersma, Pieter M. 1985. *Frisian reference grammar*. Dordrecht: Foris.
- Traugott, Elizabeth C. 1972. *A history of English syntax: A transformational approach to the history of English sentence structure*. New York: Holt, Rinehart & Winston.
- Trimingham, J. Spencer. 1946. *Sudan colloquial Arabic*. London: Oxford University Press.
- Tristram, Hildegard L.C. 1997. Do-periphrasis in contact? In *Language in time and space: studies in honor of Wolfgang Viereck on the occasion of his 60th birthday*, eds. H. Ramisch and K. Wynne, 401–17. Stuttgart: Franz Steiner.
- . 1999. *How Celtic is Standard English?* St. Petersburg: Nauka.
- , ed. 2000. *The Celtic Englishes II*. Heidelberg: C. Winter.
- Trubetskoy, Nicholas. 1931. Die phonologischen Systeme. *Travaux du Cercle Linguistique de Prague* 4:96–116.
- Trudgill, Peter. 1989. Contact and evolution in linguistic change. In *Language change: Contributions to the study of its causes*, ed. Leiv Egil Brevik and Ernst Håkon Jahr, 227–37. Berlin: Mouton de Gruyter.
- . 1996. Dialect typology: Isolation, social network and phonological structure. In *Towards a social science of language*, vol. 1: *Variation and change in language and society*, ed. Gregory Guy, Crawford Feagin, Deborah Schiffirin, and John Baugh, 3–21. Amsterdam: John Benjamins.
- . 1999. Language contact and the function of linguistic gender. *Posnan Studies in Contemporary Linguistics* 35:133–52.

- . 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5:371–74.
- Trumpp, Ernest. 1873. *Grammar of the Paštō*. Osnabrück: Biblio.
- Tryon, Darrell T., ed. 1995. Comparative Austronesian dictionary, part 1, fascicle 1. Berlin: Mouton de Gruyter.
- Tuldava, Juhan. 1994. *Estonian textbook*. Bloomington, Ind.: Research Institute for Inner Asian Studies.
- Twitchett, Denis, and John K. Fairbank, eds. 1979. *The Cambridge history of China*, vol. 3. Cambridge: Cambridge University Press.
- Ulan, Russell. 1978. Some general characteristics of interrogative systems. In *Universals of human language*, vol. 4, ed. Joseph Greenberg, 211–48. Palo Alto, Calif.: Stanford University Press.
- Upton, Clive, David Parry, and J. D. A. Widdowson, eds. 1994. *Survey of English dialects: The dictionary and grammar*. London: Routledge.
- Van den Berg, René. 1989. *A grammar of the Muna language*. Dordrecht-Holland: Foris.
- . 1996. The demise of focus and the spread of conjugated verbs in Sulawesi. In *Papers in Austronesian Linguistics*, vol. 3, ed. Hans Steinhauer, 89–114. Canberra: Pacific Linguistics.
- Van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris.
- Van Kemenade, Ans, and Nigel Vincent. 1997. *Parameters of morphosyntactic change*. Cambridge: Cambridge University Press.
- Van Klinken, Catharina Lumien. 1999. *A grammar of the Fehan dialect of Tetun*. Canberra: Pacific Linguistics (The Australian National University).
- Van Ness, Silke. 1994. Pennsylvania German. In *The Germanic languages*, ed. Ekkehard König and Johan van der Auwera, 420–38. London: Routledge.
- Vennemann, Theo. 2000. Zur Entstehung des Germanischen. *Sprachwissenschaft* 25:233–69.
- . 2001. Atlantis semitica: Structural contact features in Celtic and English. In *Historical linguistics 1999*, ed. Laurel J. Brinton, 351–69. Amsterdam: John Benjamins.
- Vergnaud, Jean-Roger, and Marie-Louise Zubizarreta. 1992. The definite determiner and the inalienable constructions in French and English. *Linguistic Inquiry* 23:595–652.
- Verheijen, J. A. J., and Charles E. Grimes. 1995. Manggarai. In *Comparative Austronesian dictionary*, part 1, fascicle 1, ed. Darrell T. Tryon, 585–92. Berlin: Mouton de Gruyter.
- Versteegh, Kees. 1984. *Pidginization and creolization: The case of Arabic*. Amsterdam: John Benjamins.
- . 1997. *The Arabic language*. New York: Columbia University Press.
- Veselinova, Ljuba. 2003. Suppletion in verb paradigms. PhD diss., University of Stockholm.
- Viereck, Wolfgang. 1993. The medieval European common market and its impact on Middle English. *Neuphilologische Mitteilungen* 94:69–78.
- Viitso, Tiit-Rein. 2003. Phonology, morphology and word formation. In *Estonian language*, ed. Mati Ereht, 9–90. Tallinn, Estonia: Estonian Academy Publishers.
- Visser, F. Th. 1963. *An historical syntax of the English language*, vol. 1. Leiden: Brill.
- Voss, Manfred. 1995. Kent and the Low Countries revisited. In *Linguistic change under contact conditions*, ed. Jacek Fisiak, 327–63. Berlin: Mouton de Gruyter.
- Voyles, Joseph B. 1992. *Early Germanic grammar*. New York: Academic.
- Wakelin, Martyn. 1972. *English dialects: An introduction*. London: Athlone.
- Wechsler, Howard J. 1979. T'ai-tung (reign 626–49) the consolidator. In *The Cambridge history of China*, vol. 3, ed. Denis Twitchett and John K. Fairbank, 188–241. Cambridge: Cambridge University Press.
- Wells, Spencer. 2002. *The journey of man*. Princeton, N.J.: Princeton University Press.



- Wells, Spencer, Nadira Yuldasheva, Ruslan Ruzibakiev, Peter A. Underhill, Irina Evseeva, Jason Blue-Smith, Li Jin, Bing Su, Ramasamy Pitchappan, Sadagopal Shanmugalakshmi, Karupiah Balakrishnan, Mark Read, Nathaniel M. Pearson, Tatiana Zerjal, Matthew T. Webster, Irakli Zholoshvili, Elena Jamarjashvili, Spartak Gambarov, Behrouz Nikbin, Ashur Dostiev, Ogonazar Aknazarov, Pierre Zalloua, Igor Tsoy, Mikhail Kitaev, Mirsaid Mirrakhimov, Ashir Chariev, and Walter F. Bodmer. 2001. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences* 98.
- Welmers, William E. 1973. *African language structures*. Berkeley: University of California Press.
- White, David L. 2002. Explaining the innovations of Middle English: What, where, and why. In *The Celtic roots of English*, ed. Markku Filppula, Juhani Klemola, and Heli Pitkänen, 153–74. Joensuu, Finland: University of Joensuu Faculty of Humanities.
- Wiesehöfer, Josef. 1996. *Ancient Persia from 550 BC to 650 AD*. London: I. B. Tauris.
- Williams-Van Klinken, Catharina, John Hajek, and Rachel Nordlinger. 2002. *Tetun Dili: A grammar of an East Timorese language*. Canberra: Pacific Linguistics (The Australian National University).
- Windfuhr, Gernot L. 1979. *Persian grammar: History and state of its study*. The Hague: Mouton.
- . 1989. New West Iranian. In *Compendium linguarum Iranicarum*, ed. Rüdiger Schmitt, 251–62. Wiesbaden: Ludwig Reichert.
- . 1987. Persian. In *The world's major languages*, ed. Bernard Comrie, 523–46. New York: Oxford University Press.
- Wittmann, Henri. 1996. Les créolismes syntaxiques du français magoua parlé aux Trois-Rivières. Communication présentée à la conférence, Avignon.
- Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. In *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, ed. Andrew B. Gonzalez, 71–91. Quezon City: Linguistic Society of the Philippines.
- . 1988. The PAN consonant system. In *Studies in Austronesian linguistics*, ed. Richard McGinn, 125–47. Athens, Ohio: Center for Southeast Asia Studies.
- . 1996. The development of the passive verb with pronominal prefix in Western Austronesian languages. In *Reconstruction, classification, description: Festschrift in honor of Isidore F. Dyen*, 15–40. Hamburg: Abera.
- Woollams, Geoff. 1996. *A grammar of Karo Batak, Sumatra*. Canberra: The Australian National University.
- Wouk, Fay. 2002. Voice in the languages of Nusa Tenggara Barat. In *The history and typology of Western Austronesian voice systems*, ed. Fay Wouk and Malcolm Ross, 285–309. Canberra: Pacific Linguistics.
- Wouk, Fay, and Malcolm Ross, eds. 2002. *The history and typology of Western Austronesian voice systems*. Canberra: Pacific Linguistics.
- Wright, Joseph. 1892. *A grammar of the dialect of Windhill in the West Riding of Yorkshire*. London: Kegan Paul.
- Wu, Yunji. 1992. *Studies in Chinese lexicology: Investigations into the Xiang dialect*. PhD diss., University of Melbourne.
- . 1999. *The development of aspectual systems in the Xiang dialect*. Paris: École des Hautes Études en Sciences Sociales.
- Wurm, Stephen A. 1977. The nature of New Guinea Pidgin. In *New Guinea area languages and language study*, vol. 3, ed. Stephen A. Wurm, 511–32. Canberra: Pacific Linguistics.
- Wurm, Stephen A., Peter Mühlhäusler, and Darrell T. Tryon, eds. 1996. *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*, vol. 2.2. Berlin: Mouton de Gruyter.
- Yar-Shater, Ehsan. 1969. *A grammar of Southern Tati dialects*. The Hague: Mouton.

- Yu, Min. 1945. Ping Robert A. D. Forrest: The Chinese Language. Review of Robert A. D. Forrest: The Chinese Language. *Yanjing Xuebao* 36:320.
- Yü, Ying-Shih. 1986. Han foreign relations. *The Cambridge history of China*, vol. 1, ed. Denis Twitchett and Michael Loewe, 377–462. Cambridge: Cambridge University Press.
- Yue-Hashimoto, Anne. 1991. Stratification in comparative dialectal grammar: A case in Southern Min. *Journal of Chinese Linguistics* 19:139–71.
- . 2003. Chinese dialects: Grammar. In *The Sino-Tibetan languages*, ed. Graham Thurgood and Randy J. LaPolla, 84–125. London: Routledge.
- Zarrinkub, ‘Abd al-Husain. 1975. The Arab conquest of Iran and its aftermath. In *Cambridge history of Iran*, vol. 4, ed. R. N. Frye, 1–56. Cambridge: Cambridge University Press.
- Zeimal, E. V. 1983. The political history of Transoxiana. In *The Cambridge history of Iran*, vol. 3[1], ed. Ehasn Yar-Shater, 232–62. Cambridge: Cambridge University Press.
- Zhou, Minglang. 1998. Tense/aspect markers in Mandarin and Xiang dialects. In *Sino-Platonic Papers* 83. Philadelphia: University of Pennsylvania, Department of Asian and Middle Eastern Studies.
- Zwettler, Michael. 1978. *The oral tradition of classical poetry: Its character and implications*. Columbus: Ohio State University Press.

*This page intentionally left blank*

# Index

- AAVE. *See* African American Vernacular English
- abbreviation, in Classical Arabic, 187–88
- “Abrupt creolization,” 253
- Abynei, 150
- Acehnese, 207–8, 234
- Achaemenids, 155–58, 163
- actualization, 10
- Adelaar, K. Alexander, 213, 219–24, 237
- adjectival concord, in Iranian languages, 148
- affixes, 122
- Afri in Cantonese, 109–10
- “drop into disuse,” 73
- in English, 73
- in Mandarin Chinese, 109–10
- plural in Iranian languages, 148
- pronominal in Tukang Besi, 203*t*
- can American Vernacular English (AAVE), 196, 258–60, 275
- Afrikaans, 70, 71, 191, 257, 260, 265–66
- be with past participles in, 75
- Dutch and, 102
- English compared to, 85, 101–2
- external possessor constructions and, 64, 67
- grammatical gender marking on the article and, 69
- inflection and, 146
- man in, 82
- second-language acquisition in, 69
- suffixes and, 72
- thou and, 81
- verb prefixes in, 71
- V2 and, 78
- agreement, in Malay, 211–12
- agglutination, 9
- Ahlgren, Arthur, 65–66
- Algonquian language, 8
- allomorphy, 6, 111–12, 115, 212*t*. *See also*
- Pronominal allomorphy
- allomorphy in verbal conjugation paradigms, 142–43
- Alpine Romansch, 3
- Altaic legacy, 104–6, 126, 130–36
- Altoma, Salih J., 167
- Alvdålska, 275
- Ambonese imprint, 247
- analytic comparative, in Arabic, 174
- analytic possessive, in Arabic, 173
- analyticity, drift to, 6–8, 15, 45, 51, 151
- Andamanese, 11
- Andi, 26
- animacy, in Slavic languages, 9–10
- Ansaldi, Umberto, 232
- Arabic, 17, 55, 57. *See also* Bedouin Arabic; Chadian Arabic; Classical Arabic; Cypriot Maronite Arabic; Gulf Arabic; Judaic Arabic; Maltese Arabic; Moroccan Arabic; Najdi Arabic; New Arabics; Nubi Creole Arabic; Sedentary Arabics
- analytic comparative in, 174
- analytic possessive in, 173

Arabic, (*continued*)

- Bedouin question of, 193–95
- bimorphemic question words in, 175
- broken plurals in, 34*t*
- business as usual in, 165–67
- case markers in, 172, 177–78
- cline of reduction in, 191–92
- clinical rather than polar, 167–68
- close shave in, 167–78
- data on, 168–69
- descendants of, 177
- diachrony/dialectology, 165
- dual marking in, 173
- features of, 169–77
- feminine comparative in, 174
- feminine singular with inanimate plurals in, 172–73
- gender in second-person/third-person plural in, 173
- genderless numbers in, 173
- inflection along cline, 192–93
- innovations in, 178–83
- interdentals in, 169
- internal passive in, 175
- introduction to, 165
- Measures in, 168, 169, 175, 177, 179
- pidgin Arabic precursors of, 195–96
- reduction in, 166
- relative marker in, 175
- short vowel reduction in, 172
- spread of, 189–93
- subjunctive/jussive in, 175–76
- SVO in, 176
- tense-blind negation in, 174
- verb patterns in, 176–77
- vowel collapse in, 172
- weak glottal stop in, 172

## Arabist literature

- case/definite marking, 183–85
- koineization in, 185–89

## Aramaic, 6, 186

- innovations in, 182
- Old Persian and, 156
- retentions of, 181–82

## Areal influence

- in language contact, 256
- in Malayo-Javanic, 240–42

## Arends, Jacques, 46

## Armenian, 139

## Aronoff, Mark, 3

## articles, 23

## aspect, 24

## aspect marking

- in Cantonese, 110–11, 118
- in Mandarin Chinese, 110–11, 117–18
- in Xiang, 117–18

## audience

- macrosociolinguistics *v.*
- microsociolinguistics, 57–58
- “real linguistics,” 55–57

## Auslautgesetze, 139

## Australian Aboriginals, 189

## Austronesian, 6, 8, 251

- ergativity in, 214–15
- evolution *v.* Malay, 212–16
- Iban trait in, 216

## Malay and, 197–200

## phonemes in, 209

## plurality in, 214

as synthetic, 281*n*8

## Avestan, 150–51

## Baagandji, 241, 242

## Baale, 15

## Baba Malay, 231–33

## Bactrian, 151, 152, 159

## Badjiri, 241

## Bailey, Charles-James, 4, 17, 59

## Balinese, 206

## Ballard, W. L., 132

Balochi, 139, 145, 148, 149, 149*t*

## Baltic languages, 8

## Bandelt, H. J., 129

## Banjar Malay, 222

## Bantu languages, 9, 22, 33, 177, 229

## Barnes, William, 92

## Bartangi, 145

## Bazaar Malay, 98, 222, 235–36

## be with past participles

- in Afrikaans, 75
- in Danish, 77
- in English, 75–77
- in Germanic languages, 75–77
- in Norwegian, 77
- in Old English, 75–76
- in Swedish, 76–77

## become-passive, 94–95

## Bedouin Arabic, 168, 180, 193–95

## case markers in, 172

## short vowel reduction in, 172

- Benjamin, Geoffrey, 226–28  
 be-perfect, 278<sup>n10</sup>  
 Berber languages, 190  
 Berbice Dutch Creole, 79, 265  
 be-verbs, 25  
 Bichlmeier, Harald, 14  
 Bickerton, Derek, 3, 18, 273  
 bimorphemic question words, in Arabic, 175  
 bioprogram, 18  
 Bisang, Walter, 131  
 Blanc, Haim, 166, 193  
 Blau, Joshua, 166, 183–85, 189  
 bleaching, 12  
 Bloch, Ariel, 188  
 Boas, Franz, 6–7, 51  
 borrowing  
   cline of in language contact, 254–55  
   Kaufman and, 255–57  
   lexical, in language contact, 255  
   Thomason and, 255–57  
 Bosman, D. B., 10  
 Breeze, Andrew, 89  
 Breton, 66  
 Brilioth, Börje, 82  
 broken plurals, in Arabic, 34<sup>t</sup>  
 Brorström, Sverker, 76  
 Brustad, Kristen E., 179  
 Brythonic Celtic, 88–90  
 Buginese, 55, 209, 215  
 Bulgarian, 96, 177  
  
 Cantonese, 49  
   affixes in, 109–10  
   aspect marking in, 110–11, 118  
   classifiers in, 109  
   complementizers in, 112–13  
   final consonants in, 107  
   grammatical gender in, 117  
   Mandarin Chinese v., 106–13  
   negator allomorphs in, 111–12  
   sentence-final particles in, 113  
   tone change in, 107–8  
   tones in, 107  
 Cape Verdean Creole, 79  
 case markers  
   in Arabic, 172, 177–78  
   in Bedouin Arabic, 172  
 case marking  
   in Pashto, 140  
   in Persian, 140, 146  
  
 Cavalli-Sforza, Francesco, 161  
 Cavalli-Sforza, Luigi Luca, 161  
 Celtic languages, 29  
 Celtic transfer, 265–66  
 Central Malayo-Polynesian  
   Tetun, 242–46  
   Timor/Flores Island, 246–49  
 Chadian Arabic, 177  
 Chao, Yuen Ren, 109, 113  
 Chappell, Hilary, 105, 106, 122  
 Chaudenson, Robert, 10, 196  
 Chauzhou Min, sandhi rules in, 114<sup>t</sup>  
 Chechen, 11–12, 13, 15  
 Chen, Chung-Yu, 135  
 Cheng, Robert L., 115  
 Chinese languages, 55. *See also*  
   Cantonese; Gan; Hakka; Mandarin  
   Chinese; Middle Chinese; Min; Old  
   Chinese; Toishan Cantonese; Wu;  
   Xiang  
   complementizers across, 120<sup>t</sup>  
   distributional anomaly in, 128–30  
 Chinook Jargon Creole, 263, 273  
 Chomsky, Noam, 57  
 Chukchi, 228  
 Classical Arabic, 165, 174, 175, 178, 272,  
   276  
   abbreviation in, 187–88  
   imperfect/perfect verbs in, 176<sup>t</sup>  
   language transmission in, 192  
   Latin/French compared to, 166  
   losses in, 169, 170<sup>t</sup>–171<sup>t</sup>  
   new Arabics and, 179, 185, 196  
   rote learning in, 168  
   subjunctive marking in, 178, 184<sup>t</sup>  
   subjunctive/jussive in, 178  
   subtraction in, 167  
   today, 186  
   verb morphology of, 180<sup>t</sup>  
 Classical Malay, 219–22, 235  
 classifiers  
   in Cantonese, 109  
   in Malay, 202, 209  
   in Mandarin Chinese, 109, 118  
   in Tukang Besi, 202  
 Cleasby, Richard, 64  
 clitics, 32–33, 47, 271<sup>t</sup>  
 Cohen, David, 169, 183, 188  
 colloquial Singapore English, 258, 262  
 COMP, 80

- complementizers, 121–22  
   in Cantonese, 112–13  
   across Chinese languages, 120*t*  
   in Hakka, 119  
   in Mandarin Chinese, 112–13, 115, 119–20  
   in Min, 115  
   in nonstandard Mandarin Chinese, 130*t*  
 complexity. *See also* Grammatical complexity  
   context as, 51–54  
   degree of in English, 60–61  
   in English, 52  
   inflection and, 35–45  
   in Iranian languages, 148–50  
   of language, 5  
   in Malay, 208*t*  
   of Malayo-Javanic, 238  
   in simplification, 9  
 conjugational classes, 31  
 conspiracy, 8  
 context  
   as complexity, 51–54  
   in Saramaccan, 49  
 contextual recovery, 53  
 copulas, 55  
   in Malay, 209–10  
   Mandarin Chinese and, 119  
   overspecification, 24–25  
 Cornips, Leonie, 63  
 Corriente, Federico, 183, 184  
 Cowan, William G., 188  
 Creole languages, 3, 5, 23, 124. *See also*  
   Berbice Dutch Creole; Cape Verdean  
   Creole; Chinook Jargon Creole;  
   Fongbe; Haitian Creole; Hawaiian  
   Creole English; Mauritian Creole;  
   Nubi Creole Arabic; Palenquero  
   Creole Spanish; São Tomense Creole  
   Portuguese; Saramaccan Creole;  
   Semicreoles; Tok Pisin  
   abnormal development in, 274  
   as full, 17  
   label of, 59  
   language contact and, 261–65  
   non-native acquisition as norm, 12  
   simplification and, 272–74  
 Creolization, 16, 79, 227, 229, 252–53  
 Crowley, Terry, 9  
 Crystal, David, 3  
   culture, grammar and, 13–14  
 Cumming, Susanna, 220  
 cumulative weightings, 68  
 Cushitic, 15  
 Cypriot Maronite Arabic, 188  
 Dahl, Ósten, 8, 257  
   on disruptive language, 14  
   on linguistic evolution, 19  
 Dalby, Andrew, 3, 197, 251, 276  
 Danchev, Andrei, 80, 88  
 Danelaw, 90–92, 98, 100–101  
 Danicism, 81  
 Danish, 71  
   be with past participles in, 77  
   grammatical gender marking on the  
     article and, 68–69  
 Dari Persian, 146  
 Darius, 160  
 Daro Batak, 208, 215  
 Darwin, Charles, 10  
 dative sickness, 66–67  
 Davidoff, Jules, 53  
 Davies, Ian, 53  
 Dawanese, 245, 245*t*  
 De Dardel, R., 270  
 De Miroschedji, Pierre, 158  
 De Rachewiltz, 126  
 De Rooij, Vincent A., 18, 260–61  
 declension, 31  
 definiteness marking, 23  
 DeGraff, Michel, 18, 253  
 demonstrative gradation  
   in Shengxian Changle Wu, 119*t*  
   in Tukang Besi, 26*t*  
   in Wu, 119  
 demonstratives  
   in Malay, 202, 209  
   in Tukang Besi, 202, 202*t*  
 Denison, David, 77  
 derivational morphology  
   English and, 69–70  
   French and, 70  
   in Germanic languages, 69–70  
   Middle English and, 70  
   Old English and, 70  
 derivational prefixes, 94  
 determiners, 201–2  
 Diem, Werner, 10, 166, 185–86, 188, 196

- directional adverbs, 92  
 English and, 74–75  
 Frisian and, 75  
 in Germanic languages, 74*t*  
 Mainland Scandinavian and, 75  
 Old English and, 74–75
- Dixon, R. M. W., 51, 136
- Dobrovolsky, Michael, 3
- Dongxiang, 133–34
- Donohue, Mark, 200, 202, 218
- do-support  
 in English, 5, 9, 60  
 source of, 89
- Dravidian language, 90
- Drift, 243–46
- drift to analyticity, 6–8, 15, 51, 151
- Dryer, Matthew, 131, 136
- dual marking, 15, 173
- Dutch, 70, 71, 103. *See also* Low Dutch  
 Afrikaans and, 102  
 English and, 95  
 grammatical gender marking on the  
 article and, 67–68  
 Limburg dialect of, 63  
 man in, 82  
 thou in, 81
- Dwyer, Arienne, 135
- Dyen, I., 198
- Eastern Abruzzese, 154
- Edwards, John, 3
- Elamites, 157–58
- Elgibali, Alaa, 195
- “elite dominance,” 161
- English, 8, 22, 23. *See also* African  
 American Vernacular English;  
 Colloquial Singapore English;  
 Hawaiian Creole English; Middle  
 English; Old English  
 abnormal development in, 274  
 affixes in, 73  
 Afrikaans compared to, 85, 101–2  
 be with past participles in, 75–77  
 be-verbs in, 25  
 Celtic transfer in, 265–66  
 complexity in, 52  
 degree of complexity in, 60–61  
 derivational morphology, 69–70  
 description v. explanation in, 72–73  
 directional adverbs and, 74–75  
 disruption in transmission of, 86  
 do-support in, 5, 9, 60  
 Dutch and, 95  
 emergence of, 59–60  
 external factors a work in, 15  
 external possessor constructions and, 64–  
 67  
 Frisian and, 95  
 Germanic languages and, 74  
 grammar of, 13  
 grammatical gender in, 31, 67–79  
 habitual marking in, 24  
 implications of losses in, 83–87  
 inflection and, 146  
 inflectional loss and, 60, 146  
 inherent reflexivity and, 61–64  
 irregular root forms, 34  
 language contact and, 257  
 losses in, 83*t*, 84*t*  
 man in, 81–83  
 modern, 57  
 Old Norse and, 97  
 passive marking with become in, 77–78  
 Scandinavian impact on, 87–101  
 suffixes in, 71–72  
 thou in, 80–81  
 V2 and, 78–80  
 Yiddish compared to, 85
- Eratosthenes, 156
- ergative languages, 26
- ergativity, 32  
 in Austronesian, 214–15  
 in Iranian languages, 148  
 overspecification, 26  
 in Pashto, 144  
 in Persian, 144
- Eskimo-Aleut, 129
- esoteric languages, exoteric language v., 11
- Estonian  
 cases in, 36*t*  
 form/semantics of, 49  
 genitive, 37–38, 47*t*  
 grammatical complexity and, 47–50  
 irregularity in, 44–45  
 overspecification in, 40–42  
 partitive, 38–45, 47*t*  
 qualitative/quantitative alterations in,  
 277*ni*



Estonian (*continued*)

- Saramaccan v., 47–50
- structural elaboration in, 42–43
- evidential marking, 24, 55
- “exaptation,” 167
- exoteric languages, esoteric languages v., 11
- external possessives, 94
- external possessor constructions
  - Afrikaans and, 64, 67
  - English and, 64–67
  - in Germanic languages, 64–67
  - in Icelandic, 64, 66, 67
  - in Middle English, 65
  - in Old English, 65
  - in Yiddish, 64
- external possessor marking, 91

Fanakalo, 157

Farizandi, 145

Faroese, 3, 63, 87, 95, 103

- grammatical gender marking on the article and, 68–69

- man in, 81

- verb prefixes in, 71

- V2 and, 79

feminine comparative, in Arabic, 174

feminine singular with inanimate plurals, in Arabic, 172–73

Ferguson, Charles, 166, 169, 188–89, 193

Fiji, 98

Fiji Hindustani, 99, 100

final consonants, 107

Finnish, 8, 36, 66

Fischer, Wolfdietrich, 169

focus marking, 27–28

Fongbe, 25, 263–64

Foreigner Talk, 18, 97

Forster, P., 129

Fortification and Treasury tablets, 156

“fourth person,” 23

Franke, Herbert, 126

French, 22, 31. *See also* Norman French; Réunionnais French

- Classical Arabic compared to, 166

- derivational morphology and, 70

- inflection cycle from Proto-Indo-

- European, 19*t*

- irregular root forms, 34

- lexical items in, 70

- man in, 82

- partitive, 43

- suppletion in, 35

Frisian, 55, 67, 70, 71

directional adverbs and, 75

English and, 95

grammatical gender marking on the article and, 67–68

thou in, 81

Friulian, 271, 271*t*

Frye, Richard N., 155

Fula, 159–60, 241

Gan, 106, 117, 122

gatekeeper, 58

gender, 31

gender in second-person/third-person

- plural, in Arabic, 173

genderless numbers, in Arabic, 173

gender-marked third-person singular, in

- Persian, 146

Genghis Khan, 105

Georgian, 229

German, 28, 52, 70. *See also* High

- German; Low German; Nonstandard

- German; Old High German;

- Pennsylvania German

Germanic languages, 8, 24, 139

be with past participles in, 75–77

derivational morphology in, 69–70

directional adverbs in, 74*t*

English and, 74

external possessor constructions in, 64–67

Inherent reflexivity in, 61–64, 61*t*losses in, 83*t*, 84*t*

man in, 81–83

passive marking with become in, 77–78

thou in, 80–81

V2 in, 32–33, 78–80

Gil, David, 97, 223–26, 229–30, 269

Gilaki, 145, 148–49

Givón, Talmy, 225

Givón's pragmatic mode, 225–26

Gould, Stephen J., 10–11

Goyette, Stéphane, 272

grammars, 4. *See also* Universal Grammar

- complexification of, 5

- culture and, 13–14

- drift to analyticity in, 6–7

- of English, 13

- in language change, 5–15
- qualitative differences among, 56, 57
- simplification of, 5–6, 8–10
- transformation interruption of, 10–11
- grammatical complexity
  - defining, 21
  - inflection and, 18–20
  - irregularity, 33–35
  - overspecification, 21–28
  - Saramaccan morphophonemics, 45–46
  - Saramaccan tone sandhi, 46–47
  - Saramaccan v. Estonian, 47–50
  - structural elaboration, 29–33
- grammatical gender, 31, 33, 91–92
  - in Cantonese, 117
  - in English, 31
  - in Gan, 117
  - in Hakka, 117
  - in Khasi, 250*t*
  - Mandarin Chinese and, 117
  - in Pashto, 140
  - in Persian, 140, 146
  - in Xiang, 117
- grammatical gender marking on the article
  - Afrikaans and, 69
  - Danish and, 68–69
  - Dutch and, 67–68
  - English and, 67–79
  - Faroese and, 68–69
  - Frisian and, 67–68
  - Mainland Scandinavian and, 67–68
  - Old English and, 67–68
  - Swedish and, 68
- grammatical morphemes, 95–97
- Greenberg, Joseph, 30, 224
- Grijns, Cornelius D., 223, 225
- Grimm's Law, 239
- "Grossreichssprache," 155
- Gulf Arabic, 168, 178, 189
- Gurani, 145, 154
- habitual marking, 24
- Haitian Creole, 22, 263–64
- Håkannsson, Gisela, 97
- Hakka, 55, 106, 122, 123
  - complementizers in, 119
  - grammatical gender in, 117
  - possessive pronouns/alienability in, 117
- Ham, William H., 46
- Han, 126
- Hanseatic League, 88
- Harding, R., 129
- Hashimoto, Mantaro, 104, 106, 126, 131–32, 134, 136
- Haspelmath, Martin, 63, 66
- Hausa, 273
- Hawaiian Creole English, 18, 262–63, 273, 275
- Hawkins, John A., 52–53, 60
- Henne, Henry, 122
- Herodotus, 156
- heterogeneous word order, 32–33
- Hezhou, 134, 136, 279*n*7
- High German, 103
- Hill, Jane, 277*m*
- Hiltunen, Risto, 70
- Hindi, 269
- Ho, Mian-Lian, 258
- Hokkien, 232–33
- Holm, John, 257, 263
- homonymy, 82
- Hope, Jonathan, 81
- Hungarian, 66
- hunter-gatherers, 13
- Hyman, Larry, 9
- Iban, 239, 242
- Iban trait, 216
- Icelandic, 6, 63, 87, 95, 103
  - as conservative outlier, 6
  - external possessor constructions in, 64, 66, 67
  - man in, 81
  - verb prefixes in, 71
- Igbo, 242
- Ijo, 79
- imperative, 212
- inalienable possession, 23, 55, 206
- indefinite man, 95
- indigenous languages, 7, 14, 51
- Indo-European languages, 8, 23, 133, 138–39
- Indo-Iranian languages, 5
- Indonesian, 97, 218*t*. *See also* Riau Indonesian
- Indonesian-type languages (IT), 198, 206, 207, 216
  - bound morphemes in, 219
  - distinctions of, 218
  - Malay structure and, 216

- Indonesian-type languages (IT) (*continued*)  
 radical simplification in, 236  
 speakers of, 221  
 verb morphology in, 216–19
- infixation, 201
- infixes, 209
- inflected interrogative pronouns, in  
 Persian, 148
- inflection  
 Afrikaans and, 146  
 along cline in Arabic, 192–93  
 complexity and, 35–45  
 English and, 146  
 as fragile, 146  
 grammatical complexity and, 18–20  
 grammatical morphemes and, 95–97  
 in Proto-Indo-European to French, 19*t*
- inflectional loss, 4, 8  
 chain-style effects of, 60  
 English and, 60  
 Kaufman and, 96  
 Thomason and, 96
- Ingush, 229
- inherent reflexive marking, 93, 94
- inherent reflexivity  
 English and, 61–64  
 in Germanic languages, 61–64, 61*t*  
 In Middle English, 62  
 in Old English, 62–63  
 in Scandinavian languages, 63
- interdentals, in Arabic, 169
- internal passive, in Arabic, 175
- intonation, pragmatics and, 54–55
- intuition, science v., 49–50
- invisible hand concept, 69
- Iranian languages, 139, 145. *See also*  
 Middle Iranian; Old Iranian  
 adjectival concord in, 148  
 complexity in, 148–50  
 ergativity in, 148  
 Persian as atypical, 150  
 Persian compared to, 147*t*  
 plural affix and, 148
- Irish Gaelic, 99
- irregular root forms, 34
- irregularity, 33–35  
 in Estonian, 44–45  
 grammatical complexity and, 33–35  
 in Saramaccan, 44–45
- Ishkashmi, 145, 146
- Isomorphism, 99
- Italian, 96, 270
- Jakarta Malay, 225
- Japanese, 23–24, 24*t*
- Jastrow, Otto, 169
- Javanese, 198, 208, 212, 214–15, 225, 238–39  
 “Javo-Sumatran hesion,” 198
- Judaic Arabic, 196
- Kamera, 244
- Kapanga, André Mwamba, 18, 260–61
- Karok, 21–22
- Kastovsky, Dieter, 89, 96
- Kaufman, Terence, 10, 59, 67, 88, 90, 93,  
 101, 103, 126, 240–41, 252–54  
 borrowing and, 255–57  
 inflectional loss and, 96  
 transfer effects and, 95
- Kaye, Alan, 166, 175, 188
- Kayne, Richard, 79
- Keller, Rudi, 69
- Kenyan Pidgin Swahili, 259–60
- Kéo, 246, 249–51
- Kerinci Malay, 56, 213, 219, 222–23,  
 281*nb*
- Kete, 261
- Khasi, 250, 250*t*
- Khmer vocalic inventory, 250*t*
- Khoi-San languages, 8, 191, 257
- Khotanese, 152
- Khwarezmian, 151–52, 159
- Kikongo, 23–24, 24*t*
- Kilpiö, M., 77
- King James Bible, 81
- Kiparsky, Paul, 80
- Kituba, 229, 261
- Kloecke, G. G., 10
- koineization, 87, 97–98  
 in Arabist literature, 185–89  
 of Ferguson, 188–89
- König, Ekkehard, 66
- Konjo, 214, 215
- Korean, 273
- Kotsinas, Ulla-Britt, 274–75
- Kroch, Anthony, 79
- Kronoby Swedish dialect, 79
- Kuala Lumpur Malay, 224
- Kubler, Cornelius C., 135
- Kune, 9

- !Kung, 11  
 Kunwinjku, 8–9, 13, 15  
 Kurdish, 139, 145–46, 150, 154  
 Kusters, Wouter, 192–93, 195–96  
 Kwa, 8–9, 242  
 Kwaio, 23*t*, 29
- LaCroix, Pierre-François, 160  
 Lahu, 28  
 Lamarre, Christine, 119, 122, 130  
 language(s). *See also specific languages*  
   commonality among, 55  
   complexity of, 5  
   Dahl on disruptive, 14  
   normal, 11–14  
   sociological dynamics and, 57  
   of Timor/Flores Island, 246  
   “top down” simplification of, 18  
   transmission of, 4  
   worldwide view of, 7–8  
 language change  
   grammars’ role in, 5–15  
   losses/mergers/collapses in, 167  
   synchrony v. sociology in, 266–67  
 language contact, 4  
   areal influence in, 256  
   bridging gap in, 253–54  
   cline of borrowing in, 254–55  
   Creole languages and, 261–65  
   English and, 257  
   as hybridization, 15–16  
   lexical borrowing in, 255  
   Malay and, 257  
   Mandarin Chinese and, 257  
   New Arabics and, 257  
   outcome of, 16  
   prediction over description, 268–69  
   reification in, 267–68  
   Romance languages and, 269–72  
   schema of results, 254–65, 254*t*  
   simplification in, 256–61  
   Sprachbund features in, 255–56  
 Lanzhou Mandarin, 134, 136  
 Lass, Roger, 60, 64, 67–69, 85, 96, 167  
 Latin, 272  
   case marking in, 10  
   Classical Arabic compared to, 166  
   nominal morphology of, 31  
 LeCoq, Pierre, 150  
 Lefebvre, Claire, 18, 264
- lexical borrowings, 95  
   in language contact, 255  
 Li, Charles N., 107  
 Li, Ying-Che, 117  
 Lightfoot, David, 79, 98–99  
 Lingala, 229  
 “lingocentricity,” 79  
 lingua franca, 9  
   as easy to grasp, 3, 197, 276  
   Malay as, 251  
 linguistic egalitarianism, 51–52  
 linguistic equilibrium, 136  
 linguistic Society of America, 51  
 Linxia Mandarin, 135  
 literacy, 57  
 lost words, 73  
 Low Dutch, 88  
 Low German, 71, 88, 278*n*8
- M17, 160–62  
 Macassarese, 209, 215, 219, 234  
 Macdonald, R. Ross, 220  
 Macrosociolinguistics,  
   microsociolinguistics v., 57–58  
 Madurese, 198, 208, 210, 213, 238–39  
 Mainland Scandinavian, 70–71, 103, 139.  
   *See also* Scandinavian impact;  
   Scandinavian languages  
   directional adverbs and, 75  
   grammatical gender marking on the  
   article and, 67–68  
   verb prefixes in, 71  
   V2 and, 78–79  
 Malay, 57, 276. *See also* Austronesian; Baba  
   Malay; Banjar Malay; Bazaar Malay;  
   Classical Malay; Indonesian-type  
   languages; Jakarta Malay; Kerinci  
   Malay; Kuala Lumpur Malay; Pidgin  
   Malay Derived dialects; Riau  
   Indonesian; Siak Malay  
   in actuality, 222–29  
   through ages, 235  
   agreement in, 211–12  
   Austronesian and, 197–200  
   Austronesian evolution v., 212–16  
   “bad,” 222  
   Classical Malay and, 219–22  
   classification of, 198  
   classifiers in, 202, 209  
   complexity, 208*t*

Malay (*continued*)

- copulas in, 209–10
- data on, 206–7
- demonstratives in, 202, 209
- determiners in, 201–2
- evolution of, 234–37
- features of, 209–12
- imperative in, 212
- inalienable possession in, 206
- incredible lightness of, 206–12
- Indonesian typology and, 199–200
- infixation in, 201
- infixes and, 209
- introduction, 197
- language contact and, 257
- as lingua franca, 251
- morphophonemics of, 199, 199*t*, 213–14
- negators in, 210
- overspecification/complexity in, 208*t*
- phonemes in, 209
- phonemic inventory in, 200–201
- phonetic evolution in, 213
- plurality in, 214
- pronominal allomorphy in, 204–5, 204*t*
- from protolanguage to, 216–22
- Proto-Malayo-Polynesian and, 219–21
- redundant concord in, 205–6
- sisters and, 207–8
- sociohistory and, 235–36
- spoken, 222, 224
- structure in IT languages, 216
- tense-aspect marking in, 202–3
- tense/aspect marking in, 211
- Tukang Besi v., 200–206
- two radically conservative, 236–37
- verb morphology in, 216–19
- Malayo-Javanic, 237–38
  - areal influence in, 240–42
  - complexity of, 238
  - simplicity in, 238–39
  - simplification in Proto-Malayo Javanic, 239–40
- Malayo-Polynesian, 197–98, 207
- Maltese Arabic, 188
- man. *See also* Indefinite man
  - in Afrikaans, 82
  - in Dutch, 82
  - in English, 81–83
  - in Faroese, 81
  - in French, 82

- in Germanic languages, 81–83
- in Icelandic, 81
- in Middle English, 82
- in Swedish, 82
- Manchus, 105, 126
- Mandarin Chinese, 5, 49, 57, 104–5, 265.
  - See also* Lanzhou Mandarin; Linxia Mandarin; Nonstandard Mandarin Chinese
  - affixes in, 109–10
  - Altaic legacy v. reduction in, 130–36
  - aspect marking in, 110–11, 117–18
  - bird's eye view of, 120–24
  - Cantonese v., 106–13
  - classifiers in, 109, 118
  - complementizers in, 112–13, 115, 119–20
  - copulas and, 119
  - distributional anomaly and, 128–30
  - final consonants in, 107
  - foreign rulers and, 125–26
  - gradient mandarin influence, 106
  - grammatical gender and, 117
  - internal developments in, 105
  - language contact and, 257
  - migrations and, 126–27
  - Min and, 113–17
  - negator allomorphs in, 111–12, 115
  - nonstandard varieties, 129–30
  - overspecification in sisters of, 121*t*
  - phonemic inventory in, 114
  - plural pronouns in, 118–19
  - possessive pronouns/alienability in, 117
  - preliminaries in, 105–6
  - register layerings in, 105–6
  - resultativity marking in, 116–17
  - sandhi rules in, 107
  - semantic overspecification in verb strings in, 116
  - sentence-final particles in, 113
  - sisters of, 120–25
  - structural abbreviation in, 124–25
  - tone change in, 107–8
  - tones in, 107, 114
- Mande, 242
- Mandinka, 23
- Mandres Albanian, 69
- Marcantonio, Angela, 167
- Marchand, Hans, 72
- markedness, 30
- Maroldt, Karl, 4, 17, 59

- Marti, Roland, 14  
 Matisoff, James A., 28, 137  
 Matthews, Stephen, 107, 111, 231  
 Mauritian Creole, 195–96  
 Mayali dialects, noun classes/agreement in, 9*t*  
 Maybrat, 24  
 Mazandarani, 148  
 McWhorter, John H., 236, 259, 273  
 Mead, Margaret, 51  
 Media Lengua, 16, 59, 230  
 Meillet, A., 155  
 Meroney, Howard M., 74  
 Miao-Yao, 104, 132  
 Michif, 59  
 Microsociolinguistics, macrosociolinguistics *v.*, 57–58  
 Middle Chinese, 114  
 Middle English, 17, 95  
     derivational morphology and, 70  
     external possessor constructions in, 65  
     inherent reflexive marking in, 94  
     inherent reflexivity in, 62  
     man in, 82  
     as “recessive,” 72  
 Middle Iranian, 151  
 Middle Persian, 151, 280*n*4, 280*n*5  
     to Modern, 152–54  
     Parthian and, 158–59  
 Min, 5, 51, 105, 106, 123. *See also* Chauzhou Min; Southern Min  
     complementizers in, 115  
     Mandarin Chinese and, 113–17  
     negator allomorphs in, 115  
     phonemic inventory in, 114  
     phonology of Proto-Min, 114  
     resultativity marking in, 116–17  
     sandhi rules in, 114–15  
     semantic overspecification in verb strings in, 116  
     tone change in, 114–15  
     tones in, 114  
 Minangkabau, 97, 208, 210, 239, 242  
*The Missing Spanish Creoles* (McWhorter), 273  
 Mitchell, Bruce, 76–77, 94  
 Mixture, 4, 229–34  
 Mohammed, 165  
 Mon-Khmer languages, 9, 207, 234, 249–51  
 Montagnais, 28  
 mood, 24  
 Moravcsik, Edith, 23  
 Moroccan Arabic, 166, 177, 180, 186, 193  
     imperfect/perfect verbs in, 176*t*  
     verb morphology of, 180*t*  
 morphology, 5, 6, 14  
 morphophonemics, 6, 29–30  
     of Malay, 199, 199*t*, 213–14  
     in Old English, 9*t*  
     in Pashto, 141–42  
     in Persian, 141–42  
     Saramaccan, 45–46  
 Moussay, Gerard, 209  
 Muna, 211, 212*t*, 214, 217  
 Munji, 139, 145–46  
 Muruwarri, 241  
 Na-Dene, 129  
 Nahuatl, 277*m*  
 Najdi Arabic, 168, 169, 192  
 Naro, Anthony J., 10  
 Native American languages, 8, 11, 17, 23, 129, 133, 228  
 Navajo, 17, 228  
 NCSL. *See* Non-hybrid Conventionalized Second-language  
 negator allomorphs  
     in Cantonese, 111–12  
     in Mandarin Chinese, 111–12, 115  
     in Min, 115  
 negators, 25, 122  
     in Malay, 210  
     in southern Min, 25*t*  
 New Arabics, 173–76, 178, 270, 272, 276, 280*n*4  
     Classical Arabic and, 179, 185, 196  
     emergence of, 190–91  
     language contact and, 257  
     “now” in, 186*t*  
 Newmeyer, Frederick J., 51  
 Ngadha, 246, 250–51  
 Nias, 214, 215  
 Niger-Congo languages, 8  
 Non-hybrid Conventionalized Second-language (NCSL), 268, 276  
 nonstandard German, 103  
 nonstandard Mandarin Chinese, 129–30, 130*t*  
 Norman French, 4, 88  
 Norman invasion, 59

- Norman, Jerry, 114, 126, 131  
 Northern Subject Rule, 89  
 Norwegian, 71  
 Nothofer, Bernd, 198  
 noun class marking, 22–23, 22*t*  
 noun classes, 33  
 Nubi Creole Arabic, 168, 191–93, 229  
 Nurse, Derek, 268  
  
 obviative marking, 23  
 Occam's razor, 236, 249  
 O'Grady, William, 3  
 Ojibwa, 11, 23  
 Old Chinese, 114, 121–23, 125, 129  
 Old English, 90, 93, 95–96, 98  
   be with past participles in, 75–76  
   derivational morphology and, 70  
   directional adverbs and, 74–75  
   external possessor constructions in, 65  
   grammatical gender marking on the  
     article and, 67–68  
   inherent reflexivity in, 62–63  
   morphological erosion in, 75  
   morphophonemics in, 9*t*  
   passive marking with become in, 77  
   suffixes in, 71–72  
   verb prefixes in, 70–71  
   V2 in, 78–79  
 Old High German, 80  
 Old Iranian, 150–51  
 Old Norse, 68, 75, 77, 80, 90, 93–96, 99  
   be-perfect in, 278*m10*  
   English and, 97  
 Old Persian, 138–39, 141, 153–54  
   abbreviated Persian and, 150–51  
   Aramaic and, 156  
   decay of, 155  
   Elamites and, 157–58  
   in Iran, 156–57  
 O'Neil, Wayne, 91, 96  
 Oppenheimer, Stephen, 129  
 Optimality Theory, 56  
 Ormuri, 145–46, 149  
 Ossetic, 145, 148  
 Ostrobothnian Swedish, 69  
 OV word order, 60  
 overspecification  
   aspect, 24  
   copulas, 24–25  
   definiteness marking, 23  
   degrees of, 22  
   demonstrative gradation, 26  
   ergativity, 26  
   in Estonian, 40–42  
   evidential marking, 24  
   focus marking, 27–28  
   grammatical complexity and, 21–28  
   inalienable possession, 23  
   in Malay, 208*t*  
   mood, 24  
   negators, 25  
   noun class marking, 22  
   obviative marking, 22  
   person/number marking, 23  
   pragmatic particles, 28  
   relativization, 28  
   in Saramaccan, 40–42  
   semantic, in Mandarin Chinese, 116  
   in sisters of Mandarin Chinese, 121*t*  
   tense, 23–24  
   valence adjustment, 26–27  
  
 Pagel, Karl, 88  
 Palenquero Creole Spanish, 227, 259  
 Panare, 25  
 Parkvall, Mikael, 236  
 Parry, David, 91–93  
 Parthian, 151–52, 158–60  
 Pashto, 56, 139, 145–46, 148–49  
   allomorphy in verbal conjugation  
     paradigms and, 142–43  
   case marking in, 140  
   declension of Pox in, 142*t*  
   ergativity in, 144  
   grammatical gender in, 140  
   morphophonemics in, 141–42  
   Persian and, 140–45  
   plural marking in, 140–41, 141*t*  
   pronominals in, 142  
   verb conjugations in, 143*t*  
   verbal categories in, 143–44, 144*t*  
 passive marking with become, 77–78  
 Peitsara, Kirsti, 62–63  
 Pennsylvania German, 70  
 Perkins, Revere D., 11, 13, 277*n1*  
 Persian, 5, 56, 57. *See also* Dari Persian;  
   Iranian languages; Middle Persian;  
   Old Persian  
   as abbreviated morphologically, 161  
   under Achaemenids, 155–58, 163

- allomorphy in verbal conjugation  
   paradigms and, 142–43  
 anomalous nature of, 163  
 appeal to accent in, 139  
 as atypical Iranian language, 150  
 case marking in, 140, 146  
 chronological perspective of, 150–54  
 in Eastern states, 159–60  
 Eastern v. Western, 151–52  
 ergativity in, 144  
 features of, 145–50  
 gender-marked third-person singular in, 146  
 genetic evidence for, 160–62  
 grammatical gender in, 140, 146  
 inflected interrogative pronouns in, 148  
 Iranian languages compared to, 147*t*  
 as “marvelously simple,” 138–39  
 Middle to Modern, 152–54  
 morphophonemics in, 141–42  
 Pashto and, 140–45  
 plural marking in, 140–41  
 plural with numerals in, 148  
 pronominals in, 142  
 reduction in, 154–55  
 sisters of, 139  
 subtractive changes in, 153  
 verb conjugations in, 142*t*  
 verbal categories in, 143–44, 143*t*
- Persis, 156
- person/number marking, 23
- Philippines-type languages, 198
- phonemes, 209
- phonemic inventory, 30–31, 55, 277*m1*  
   in Malay, 200–201  
   in Mandarin Chinese, 114  
   in Min, 114  
   in Tukang Besi, 200–201
- phonetic erosion, 56, 64
- Pidgin, 4–5  
   Arabic precursors, 195–96  
   fallacy, 17–18
- Pidgin Hindustani, 98
- Pidgin Malay Derived dialects (PMD),  
   222–23, 227–28, 231, 233  
   big picture of, 233–34  
   sociohistory and, 224–25
- Pidgin/creole detour, 252–53
- pidginization, 79, 227, 229, 252–53
- Pienemann, Manfred, 97
- Pirahã, 268
- Platt, John T., 258
- plural affix, Iranian languages and, 148
- plural marking, 33–34  
   in Pashto, 140–41, 141*t*  
   in Persian, 140–41
- plural pronouns  
   in Mandarin Chinese, 118–19  
   in Wu, 119  
   in Xiang, 119
- plural with numerals, in Persian, 148
- plurality, 214
- PMD. *See* Pidgin Malay Derived dialects
- Poedjosoedarmo, Gloria R., 239
- Polish, 15, 167
- Polome, Edgar C., 18
- polysynthetic languages, 277*m1*
- Popular Brazilian Portuguese, 258–59
- Portuguese, 270. *See also* Popular Brazilian Portuguese; São Tomense Creole Portuguese  
   indirect/third person direct object clitic in, 271*t*  
   third-person clitic in, 271*t*
- possessive pronouns/alienability, 117
- Poussa, Patricia, 100
- pragmatic particles, 28
- pragmatics, intonation and, 54–55
- Prentice, D. J., 222–24, 237
- Principles and Parameters framework, 80
- printing, 6
- pronominal allomorphy  
   in Malay, 204–5, 204*t*  
   in Tukang Besi, 204–5, 205*t*
- pronominals, 142
- pronouns, in Kwaio, 23*t*
- Proto-Germanic features, 75, 77–78, 80–81, 85*t*–86*t*, 87, 103
- Proto-Indo-European influence, 80, 167
- Proto-Malayo-Polynesian  
   Malay and, 219–21  
   verb morphology in, 217*t*
- Proto-Malayo Javanic, 239–40
- Proto-Sino-Tibetan, 131
- Ramat, Paolo, 85
- “real linguistics,” 55–57
- reduction  
   Altaic legacy v., 130–36  
   in Arabic, 166



- reduction (*continued*)  
 cline of in Arabic, 191–92  
 mixture v. in Riau Indonesian, 229–34  
 in Persian, 154–55  
 problem of, 3–4  
 response to, 4–5  
 Scandinavian impact, 95–97  
 redundant concord, 205–6  
 reflexivity, 12. *See also* Inherent reflexivity  
 relative marker, in Arabic, 175  
 relativization, 28  
 Renfrew, Colin, 161  
 resultativity marking, 116–17  
 Réunionnais French, 190, 257–58, 260, 265, 275  
 Rhodes, Richard, 268  
 Riau Indonesian, 98, 223–24, 225  
 data for, 228  
 Givón's pragmatic mode, 225–26  
 mixture v. reduction in, 229–34  
 natural development in, 226–28  
 Sprachbund features, 226  
 theoretical inconsistency and, 228–29  
 Rissanen, Matti, 82  
 Roberson, Debi, 53  
 Roberts, Ian, 78  
 Robinson, Ian, 57  
 Rohrbacher, Bernhard, 78, 79  
 Romance languages, 6, 8, 24  
 language contact and, 269–72  
 as NCSLs, 269  
 Romanian, 67, 269, 281*m*  
*The Roots of Language* (Bickerton), 273  
 Rosenhouse, Judith, 175, 194  
 Roshani, 149, 150  
 Ross, Malcolm, 217, 219  
 Rotinese, 244  
 Russenorsk, 18  
 Russian, 87, 167, 269  
 Rusyn, 96  
 Rydén, Mats, 76
- Sagart, Laurent, 122  
 Saka, 151  
 Samoyedic, 167  
 Sandhi rules, 5, 9  
 in Chauzhou Min, 114*t*  
 in Mandarin Chinese, 107  
 in Min, 114–15  
 Toba Batak in, 214  
 tone, in Saramaccan, 46–47
- Sangesari, 145  
 São Tomense Creole Portuguese, 273  
 Sapir, Edward, 15, 75  
 Sapir-Whorf hypothesis, 13, 53  
 Saramaccan, 12, 13, 35–36, 48, 264  
 context in, 49  
 Estonian v., 47–50  
 genitive/partitive marking in, 47*t*  
 grammatical complexity and, 45–50  
 irregularity in, 44–45  
 morphophonemics, 45–46  
 overspecification in, 40–42  
 structural elaboration in, 42–43  
 tone sandhi, 46–47  
 Saramaccan Creole, 34  
 Sasak, 211  
 Sawyer, P. H., 101  
 Sayehli, S., 97  
 Scandinavian impact, 87  
 be-perfect and, 278*m*10  
 Brythonic Celtic, 88–90  
 on English, 87–101  
 evidence for, 100–101  
 Low Dutch, 88  
 Norman French, 88  
 process of elimination, 90–91  
 reduction, 95–97  
 simplification in, 98–100  
 timing, 91–93  
 transfer, 93–94  
 underspecification in, 98–100  
 Scandinavian languages, 59, 63. *See also*  
 Mainland Scandinavian
- Scherre, Maria Mara Pereira, 10  
 science, intuition v., 49–50  
 Sedentary Arabics  
 innovations in, 178–80  
 perspective on, 180–81  
 Selcan, Zülfü, 150  
 semantic overspecification in verb strings, 116  
 semicreoles, 257, 266, 274  
 Semitic languages, 8  
 Aramaic, 181–82  
 mutation of, 181–83  
 numbers of, 185  
 South Arabian, 182–83  
 Semitophone Fertile Crescent, 161

- Semnani, 145  
 sentence-final particles, 113  
 Serbo-Croatian, 32, 96, 124, 167  
 Shaba Swahili, 18, 259–61  
 Shengxian Changle Wu, 197  
 short vowel reduction, 172  
 Shugni, 139, 145–46, 149–50  
 Siak Malay, 228  
 Siegel, Jeff, 97, 262  
 Sika, 244  
   “eat” in, 245*t*  
   verb paradigms in, 244*t*  
 simplification  
   as background factor, 15–17  
   Boas on, 6–7  
   complexity in, 9  
   counterbalancing of, 8–10  
   Creole languages and, 272–74  
   of grammars, 5–6, 8–10  
   in language contact, 256–61  
   in Proto-Malayo Javanic, 239–40  
   in Scandinavian impact, 98–100  
   “top down,” 18  
 Singlish, 258  
 “Sinosphere,” 137  
 Sivandi, 145  
 Skjaervø, P. Oktor, 145  
 Slavic languages, 8–10, 14, 177, 187  
 Smith, Henry, 67  
 Smith-Hefner, Nancy J., 225  
 Sneddon, James Neil, 200  
 Sogdian, 151–52, 159  
 Sorace, Antonella, 76  
 sound changes, 95  
 Southern Min, 25*t*, 114–16  
 Spanish, 31, 96. *See also* Palenquero  
   Creole Spanish  
 sprachbund features, 8, 63, 136  
   in language contact, 255–56  
   Riau Indonesian, 226  
 Standard Average European language, 7  
 Strang, Barbara M. H., 68, 81, 83  
 structural elaboration  
   in Estonian, 42–43  
   grammatical complexity and, 29–33  
   in Saramaccan, 42–43  
 subjunctive/jussive  
   in Arabic, 175–76  
   in Classical Arabic, 178  
 Sudanese, 55, 215  
 suffixes, 71–72  
 Summer Institute of Linguistic  
   Ethnologue, 206  
 Sundanese, 198, 208, 238–39  
 Sundermann, Werner, 155  
 Suppletion, 35  
 SVO, 79, 176  
 Swahili, 22*t*, 33, 35–36, 259. *See also*  
   Kenyan Pidgin Swahili; Shaba Swahili  
 Swedish, 71, 274–75. *See also* Kronoby  
   Swedish dialect; Ostrobothnian  
   Swedish  
   be with past participles in, 76–77  
   grammatical gender marking on the  
   article and, 68  
   man in, 82  
   passive marking with become and, 77–78  
 Sye, 268  
 sympathetic dative, 64  
 synchrony v. sociology, 266–67  
 Szemerényi, O., 156  
 Tadmor, Uri, 220  
 Tagalog, 27, 216, 268  
 Tai, 104, 132  
 Talishi, 145  
 Tamian Latvian, 69  
 Tang dynasty, 105, 127  
 Tanwang, 133–34, 136  
 Tati, 139, 145–46, 149, 154  
 Taylor, Ann, 79  
 TD. *See* Tetun Dili  
 tense, 23–24  
 tense-aspect marking  
   in Malay, 202–3, 211  
   in Tukang Besi, 202–3  
 tense-blind negation, in Arabic, 174  
 Tetun Dili (TD), 242, 249  
   drift in, 243–46  
   Tetun Terik v., 243, 251  
   Timor/Flores Island and, 247–48  
 Tetun Terik (TT), 242  
   drift in, 243–46  
   Tetun Dili v., 243, 251  
 thence, 278*n*3  
 Thomason, Sarah J., 10, 59, 67, 88, 90, 93,  
   101, 103, 126, 195, 240–41, 252–54  
   borrowing and, 255–57

- Thomason, Sarah J. (*continued*)  
 inflectional loss and, 96  
 transfer effects and, 95
- Thompson, Sandra A., 107, 109
- thou, 100  
 Afrikaans and, 81  
 in Dutch, 81  
 in English, 80–81  
 in Frisian, 81  
 in Germanic languages, 80–81
- Thurgood, Graham, 207, 231
- Thurston, William R., 13
- Timor/Flores Island  
 beyond TD/Bima-Sumba, 247–48  
 languages of, 246, 248–49  
 world languages and, 249–51
- Toba Batak, 208, 212, 214
- Toishan Cantonese, 279*n*3
- Tok Pisin, 18, 236, 264–65
- Tokodede, 249
- tone change, 122  
 in Cantonese, 107–8  
 in Mandarin Chinese, 107–8  
 in Min, 114–15
- “Tone fetish,” 120
- tones, 121  
 in Cantonese, 107  
 in Mandarin Chinese, 107, 114  
 in Min, 114
- Torrioni, A., 129
- transitivity, 63–64
- trigger weakening, 98–99
- trimming of overspecified/complex  
 features, 99–100
- Trubetskoy, Nicholas, 30
- Trudgill, Peter, 69, 103
- Tsez, 26, 30–31, 31*t*, 229
- TT. *See* Tetun Terik
- Tukang Besi, 214, 217, 218, 219  
 classifiers in, 202  
 demonstrative gradation in, 26*t*  
 demonstratives in, 202, 202*t*  
 determiners in, 201–2  
 imperative in, 212  
 inalienable possession in, 206  
 infixation in, 201  
 Malay v., 200–206  
 phonemic inventory in, 200–201  
 pronominal affixes in, 203*t*  
 pronominal allomorphy in, 204–5, 205*t*  
 realis/irrealis pronominal affixes in, 203*t*  
 redundant concord in, 205–6  
 tense-aspect marking in, 202–3
- Tuyuca, 24
- T-V pronoun, 81
- Twi, 9, 49, 51
- Twitchett, Denis, 126
- Tyndale, William, 62
- Ubangian languages, 261
- Uma, 213, 214
- underspecification, in Scandinavian  
 impact, 98–100
- Universal Grammar, 21
- Upton, Clive, 91–93
- Ura, 268
- Uralic languages, 8, 229
- valence adjustment, 26–27
- Vanuatu Oceanic language, 9
- Vennemann, Theo, 88
- verb conjugations  
 in Pashto, 143*t*  
 in Persian, 142*t*
- verb morphology  
 of Classical Arabic, 180*t*  
 in hypothetical IT language, 217*t*  
 in Indonesian, 218*t*  
 in IT languages, 216–19  
 in Malay, 216–19  
 in Proto-Malayo-Polynesian, 217*t*
- verb patterns, in Arabic, 176–77
- verb prefixes, 70–71
- verbal categories  
 in Pashto, 143–44, 144*t*  
 in Persian, 143–44, 143*t*
- Vergnaud, Jean-Roger, 64
- Versteegh, Kees, 4, 17, 166–67, 169, 184–85, 191, 193–95
- Viereck, Wolfgang, 88
- Vigfusson, Gudbrand, 64
- Visser, F. Th., 63–66, 72–73
- voiced initials, 121
- von Humboldt, Wilhelm, 51
- Voss, Manfred, 88
- vowel collapse, in Arabic, 172
- Voyles, Joseph B., 85
- V2, 60, 92  
 Afrikaans and, 78  
 disappearance of, 100

- English and, 78–80  
 Faroese and, 79  
 in Germanic languages, 32–33, 78–80  
 Mainland Scandinavian and, 78–79  
 in Old English, 78–79  
 triggers for, 79
- Wackernagel's Law, 188  
 Wakelin, Martyn, 93  
 Wakhi, 145, 146, 150  
 Wangkumara, 241  
 Weak glottal stop, in Arabic, 172  
 Wells, Spencer, 161–62  
 Welsh, 29, 66  
 West Atlantic languages, 8  
 White, David L., 90  
 Widdowson, J. D. A., 91–93  
 Windfuhr, Gernot L., 143, 145  
 Wolff, John, 268  
 Wolof, 241  
 Wu, 5, 106, 122–23. *See also* Shengxian  
     Changle Wu  
         demonstrative gradation in, 119  
         plural pronouns in, 119  
 Wüest, J., 270
- Wutun, 134, 136  
 Wycliffe, John, 62
- Xiang, 106, 122  
     aspect marking in, 117–18  
     grammatical gender in, 117  
     plural pronouns in, 119
- Yaghnobi, 145, 151  
 Yazgulami, 146, 149–50  
 Yiddish, 70, 75, 103  
     English compared to, 85  
     external possessor constructions in,  
         64  
 Yidgha, 146  
 Yip, Virginia, 107, 111  
 Yoruba, 9, 242, 259  
 Yue, 106  
 Yue-Hashimoto, Anne, 118, 119, 131  
 Yupik, 11, 27
- Zaza, 145, 150, 154  
 Zhou, Minglang, 117–18  
 Zubizarreta, Marie-Louise, 64  
 Zwettler, Michael, 10, 183–84