

Copyright
by
Alexander Magidow
2013



The Dissertation Committee for Alexander Magidow
certifies that this is the approved version of the following dissertation:

**Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic
Dialect Diversity**

Committee:

Kristen Brustad, Supervisor

Patience Epps, Co-Supervisor

Mahmoud Al-Batal

Na'ama Pat-El

Mohammad A. Mohammad

**Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic
Dialect Diversity**

by

Alexander Magidow, B.A., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2013

Dedicated to the curiosity which drives us to question how things came to be.

Acknowledgments

This dissertation would not have been possible without my amazing committee, especially my co-supervisors Kristen Brustad and Pattie Epps. They gave me an incredible amount of support and encouragement, allowing me to dream while keeping me grounded. I am immensely grateful to the entire committee for their expertise, and for their attention both to broad issues and to small mistakes. All remaining mistakes are solely the responsibility of the author.

The support of my colleagues was a great help, especially when working through a difficult point. I would like to acknowledge in particular the members of UT's Center for Arabic Dialect Research (CADR) and the Multidisciplinary Approaches to Linguistics (MAL) graduate presentation group, where I was able to get feedback on many of the ideas that made it to the dissertation, and many that did not. Thanks also go to the audiences at the Jil Jadid conferences and the Arabic Linguistic Symposium.

My wife Melanie was the touchstone that kept me sane throughout this project and reminded me that there is so much to life beyond the dissertation.

This dissertation is very much indebted to the open source and free software movements. The dissertation was compiled in \LaTeX using the \TeX distribution and the \TeX Maker editor. Most of the figures were made using the Inkscape vector graphics program. The Roman text was set in Linux Libertine,

and the Arabic text was set in Scheherazade from SIL. Since the figures are complex, the reader is encouraged to use the most recent version of their PDF reading software.

The dissertation and all of the original figures are licensed under the Creative Commons ‘Attribution-NonCommercial-NoDerivs’ license, which allows for free distribution of the dissertation or figures for non-commercial purposes provided they are attributed to the author.

Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity

Publication No. _____

Alexander Magidow, Ph.D.
The University of Texas at Austin, 2013

Supervisors: Kristen Brustad
Patience Epps

This dissertation develops a new framework for reconstructing the diversity of a language at a given historical time period. It applies this framework to the problem of reconstructing the diversity of Arabic dialects immediately prior to the Islamic conquests, which spread speakers of these dialects across much of North Africa and the Middle East. The study first establishes a theoretical framework for reconstructing historical speech communities, defined as groups of speakers linked by shared allegiance. It then analyzes the tribal and non-tribal social organization in Pre-Islamic Arabia, and provides a detailed historical overview of how the Islamic conquests contributed to the Arabization of the conquered territories. Finally, the dissertation reconstructs the linguistic history of the Arabic demonstratives, using them as a variable to determine which speech communities existed in pre-Islamic Arabic, where they were located in time and space, and how the diversity of those communities is related to the diversity of modern Arabic dialects.

Table of Contents

Acknowledgments	v
Abstract	vii
List of Tables	xiv
List of Figures	xv
Introduction	1
0.1 Seeking Diversity in Arabic's Past	4
0.1.1 Previous Approaches	4
0.1.2 Current Approach	7
0.1.2.1 Organization	9
0.2 Demonstratives	10
0.3 Data Sources	13
0.4 Sample points	16
0.4.1 Classical Arabic	16
0.4.2 Peripheral Dialects	17
0.4.3 North African Dialects	20
0.4.4 Nile Valley and Southern African Dialects	23
0.4.5 Levantine Dialects	25
0.4.6 Mesopotamian Dialects	27
0.4.7 Arabian Dialects	28
0.4.8 Yemeni Dialects	29

0.4.9	Summary Chart and Map of Sample Points	30
0.5	Conventions and Abbreviations	34
Chapter 1. From the genealogy of languages to the archeology of speech communities		36
1.1	Challenges in applying the comparative method	38
1.1.1	Units of analysis	38
1.1.2	Genealogy	45
1.1.2.1	Homogeneity of the past	49
1.2	From languages to speech communities	52
1.2.1	Defining speech communities	54
1.2.1.1	Social Networks	54
1.2.1.2	Identity and Allegiance	59
1.2.1.3	Speech Communities: Networks, Allegiance, Size	66
1.2.2	Uniting the Linguistic and Extra-Linguistic	70
1.2.2.1	Changes in Structure and Allegiance	71
1.2.2.2	Diagnosing Speech Communities	75
1.3	Towards an archaeology of speech communities	77
1.3.1	Alignment Events and Innovation	78
1.3.2	Diffusion within closely related lects	82
1.3.2.1	Social diffusion	83
1.3.2.2	Geographical Diffusion	86
1.3.3	Diffusion between distantly related lects	89
1.3.3.1	Degree of difference	91
1.3.3.2	Dominance	92
1.3.3.3	Integration	95
1.3.3.4	Multilingualism	96
1.3.4	Newly established communities	99
1.3.4.1	Dialect Mixing to Focusing	101
1.3.5	Shift to another lect	106
1.3.5.1	Direction of shift	106
1.3.5.2	Linguistic outcomes of shift	108
1.4	Summary	110

Chapter 2. Pre-Islamic Arabia: Tribes and Social Organization	112
2.1 Modes of Subsistence	113
2.2 Group Organization	119
2.3 Group Interactions	122
2.4 The Tribe as a Linguistic Unit	127
2.5 The Bedouin Paradox	133
2.6 Conclusion	136
 Chapter 3. History: Greater Arabian Peninsula	 138
3.1 Approach	140
3.2 General Trends	143
3.2.1 Subsistence Patterns and Barriers	143
3.2.2 Linguistic Separation	147
3.2.2.1 New Cities (<i>ʿamṣār</i>)	148
3.3 Arabian Peninsula	150
3.3.1 Ancient North Arabian and Arabic	150
3.3.2 Population movements in the Arabian Peninsula prior to Islam	156
3.3.3 Yemen	160
3.3.4 Oman	164
3.3.5 Summary: Arabian Peninsula	166
3.4 The Levant	166
3.5 Pre-Islamic Linguistic Situation	167
3.5.1 Client States and the Arabization of the South-central Levant	169
3.5.2 From Roman to Arab Rule	173
3.5.3 Urban settlement	175
3.5.4 Natural Disasters: Earthquake and Plague	179
3.5.5 The Lebanese Mountains	184
3.5.6 Summary: Arabization of the Levant	185
3.6 Mesopotamia	187
3.6.1 Pre-Islamic Linguistic Situation	188
3.6.2 Conquest and Continuity	192
3.6.3 New-cities founded in Mesopotamia	199

3.6.3.1	Kufa	199
3.6.3.2	Basra	204
3.6.3.3	Baghdad	205
3.6.4	Other settlements and settlement patterns in Mesopotamia	206
3.6.5	Summary: Arabization of Iraq	209
3.7	Summary: Arabic and the Greater Peninsula	210
Chapter 4.	History: Outside the Greater Peninsula	212
4.1	Egypt	212
4.1.1	Settlement of Egypt	214
4.1.2	Arabization of Egypt	219
4.1.3	Summary: Arabization of Egypt	222
4.2	North Africa	223
4.2.1	Westward Migrations	233
4.2.1.1	Historical Overview of Banū Hilāl	237
4.2.1.2	Re-evaluating Origins	240
4.2.1.3	Re-evaluating Timing	241
4.2.1.4	Re-evaluating Impact	243
4.2.1.5	Western Migrations Reconsidered	246
4.2.2	Effects of the Reconquista on North Africa	247
4.3	Andalusia	251
4.4	Peripheral regions	257
4.4.1	Malta and Sicily	257
4.4.2	Cyprus	260
4.4.3	Central Asia	261
4.4.4	Sudan, Chad, Nigeria	262
4.5	Conclusion	262
Chapter 5.	The Linguistic Development and Reconstruction of Arabic Demonstratives	265
5.1	Typology of Demonstratives	266
5.1.1	Arabic Demonstratives from a Typological Perspective	273
5.2	Demonstratives in Semitic	278

5.2.1	Functions of Semitic Demonstratives	286
5.2.2	Semitic and Arabic demonstrative development	288
5.3	Demonstratives derived from * <i>ḏa</i> :	292
5.3.1	Development of singulars	293
5.3.1.1	Gender differentiation	293
5.3.1.2	Diminutives	301
5.3.2	Development of plurals	304
5.3.2.1	Plural forms with /l/	304
5.3.2.2	Other Plural Forms	309
5.4	The evolution of * <i>ha</i> :	312
5.4.1	Original function of * <i>ha</i> :	312
5.4.2	* <i>ha</i> : as a presentative	314
5.4.2.1	Attachment to <i>ḏa</i> : forms	317
5.4.3	* <i>ha</i> : as a demonstrative	332
5.4.3.1	* <i>ha</i> : as an adnominal demonstrative	333
5.4.3.2	* <i>ha</i> : as a pronominal demonstrative	336
5.4.4	Elaborated * <i>ha</i> : forms	338
5.4.4.1	Elaborated * <i>ha</i> : forms with * <i>ḏa</i> :	340
5.4.5	Word-final - <i>ha</i> :	346
5.5	Linguistic Chronology of Arabic Demonstratives	353
5.6	Conclusion	357
Chapter 6.	Social, Linguistic, Historical Analysis and Reconstruction	358
6.1	Consonant-Alternating Dialects	359
6.2	<i>Hawla</i> dialects	365
6.2.1	<i>hawla</i> dialects in the Levant	369
6.2.2	<i>hawla</i> dialects in Mesopotamia	375
6.2.3	<i>ha:k(a)</i> developments	376
6.3	Dialects with <i>ḏawla</i> , <i>ḏayla</i> plurals	382
6.3.1	Free variation of <i>ḏawla</i> and <i>ḏayla</i>	384
6.3.2	Gendered <i>ḏawla</i> , <i>ḏayla</i>	384
6.3.3	<i>ha:ḏo:l</i> and <i>ha:ḏo:l</i> , <i>ha:ḏanna</i> dialects	388

6.4	Dialects with <i>ha:ðu:</i> plurals	396
6.5	Maltese Arabic	402
6.6	Classical Arabic	405
6.7	Dialect Classification	409
Conclusion		421
6.8	Evaluating the Speech Communities Approach	424
6.8.1	Tribes as linguistic units	427
6.8.2	Semitics and Arabic	428
6.9	Out of Yemen or Out of Levant	430
6.10	Successes and Failures of Arabic	437
6.11	Further Research	441
Appendix		443
Bibliography		450

List of Tables

0.1	Summary of Dialect Abbreviations	30
0.2	List of conventions	35
3.1	Plagues in the Levant	180
5.1	Pronominal and adnominal demonstrative forms from CHDSH. . .	274
5.2	Classical Arabic dual demonstratives	278
5.3	Common Semitic (Proximal) Demonstratives	279
5.4	Common West Semitic (Proximal) Demonstratives	279
5.5	Summary of Semitic Demonstrative Formatives	286
5.6	Potential sites for analogous extension of gender and number marking	295
5.7	Demonstrative forms in Yemen #130.	301
5.8	Plural demonstrative forms derived from <i>*ʔVl(a)</i>	306
5.9	Proximal demonstratives with final /n/-sounds in plural	309
5.10	<i>hawd</i> -type demonstrative paradigms	341
5.11	Anatolian dialects with <i>ha:k</i> demonstratives	342
5.12	Stages in the development of <i>hawd</i> -dialects	344
5.13	EGCAI Demonstratives	346
5.14	NIG Demonstratives and Presentatives	350
5.15	Early Arabic dialectal demonstratives	354
6.1	Demonstratives in Yemen #120	377
A.1	Demonstrative data from sample: Proximal forms	444
A.2	Demonstrative data from sample: Distal forms	447

List of Figures

0.1	Map of Sample Dialects	33
1.1	Two connected networks	57
1.2	Divergence along the German-Netherlands Border	74
2.1	Schematic Map of Arabian Agricultural Areas	115
3.1	Map of the Greater Arabian Peninsula	139
3.2	Map of Rainfall in the Middle East	144
3.3	Map of speech in Yemen according to al-Hamdani.	165
3.4	Communities in Lebanon 850–1150	186
3.5	Conversion to Islam in Iraq	197
3.6	Conversion to Islam in Iraq among Iraqis and Persians	198
3.7	Reconstructed plan of Kufa	200
4.1	Conquests of North Africa	226
4.2	North African dominions in the ninth century	231
4.3	Islamic conversion curves from Egypt and Tunisia	231
4.4	Original homelands of Banū Hilāl and Banū Sulaym	236
4.5	Diagram of Arab gene markers in North Africa	245
4.6	Map of major Islamic and post-Islamic movements	264
5.1	Map of Yemeni Proximal Singular Demonstratives	297
6.1	Plural demonstrative forms in Yemen	367
6.2	Dialects Syria by Stage of Development	373

6.3	Map of Demonstratives in Mesopotamia	379
6.4	Schematic of Arabic dialect development with respect to demonstratives.	418
6.5	Schematic of Pre-Islamic Speech Communities	420

Introduction

The modern Arabic dialects present a considerable challenge for the historical linguist. They are incredibly diverse and are spoken from Mauritania to Afghanistan, bringing them into contact with each other and with a variety of other languages. Dialects can differ significantly even within a small area, and many of the apparently minor linguistic differences between dialects have taken on the status of shibboleths, so they have become charged with sociolinguistic meaning. The fact that many speakers of Arabic dialects practiced a transhumanist lifestyle until the modern era also means that some linguistic forms have spread more widely than they might have among a purely sedentary population. Few clear national or even areal standard dialects have developed, and so historical linguists are faced with the difficulty of determining what dialects should be included in their samples. At the same time, Classical Arabic, the product of the standardization of various pre-Islamic literary registers, has an unclear relationship to the modern Arabic dialects — though it seems to preserve archaic features not attested in modern dialects, these features could actually be innovations of the literary register, and thus never a part of earlier dialects.

Furthermore, the most important single historical event with regard to the spread of Arabic is not its initial pre-historic dispersion from its proto-language community, but rather the historic rise of the Arabic-speaking Islamic caliphates in

the seventh century CE. Arabs as a separate social and linguistic group are attested at the very beginning of the first millennium BCE, which means that there was over a millennium and a half at least for the Arabic dialects to differentiate prior to the Islamic conquests and expansion. Therefore, to understand the modern diversity of Arabic dialects, the historical linguistic must have a goal beyond reconstructing proto-Arabic.

Many studies on the history of Arabic have been limited by the focus on Classical Arabic as the origin of the Arabic dialects. It was felt that the transition from Classical Arabic to the modern dialect could only be explained by some sort of radical sociolinguistic event. Ferguson (1959) suggested that colloquial dialects arose from a ‘military camp koine’ during the time of the conquests, while Versteegh (1984) proposed a process of pidginization and decreolization to account for these apparently radical differences between Classical Arabic and modern Arabic dialects (these approaches are summarized and critiqued in A. M. Miller, 1986).¹ The most recent study in this vein is al-Sharkawi (2010), who attempts to explain the differences between Classical Arabic and (primarily modern Egyptian) colloquial Arabic as a result of the use and conventionalization of foreigner talk.

The focus on Classical Arabic has also spawned a literature devoted to understanding the diversity of Classical Arabic, as reported in the Classical grammatical tradition, with varying quality of work. Fück (1950) is essentially a translation of various Arab grammarians’ work that dealt with diversity in Classical Arabic. Though that work made those sources accessible to a generation of researchers unable to work with the source texts, it did not necessarily synthesizing those works in a meaningful way.

¹ See also al-Sharkawi (2010, Chapter two), Al-Jallad (2012, Chapter one) and Abboud-Haggag (*EALL: “Dialects: Genesis”*) for further summaries.

Rabin (1951) is a very comprehensive attempt to synthesize the various reports of variations in Classical Arabic and assign those variants to tribes living in the pre-Islamic world. The primary weakness of this text is that Rabin does little to distinguish between grammatical treatises produced in the eighth century and the fifteenth, relying primarily on the latter, and it is hard to take such late accounts at face value.² Al-Jundī (1983) undertakes a very similar project of cataloging Classical Arabic dialect diversity, and is only slightly more conservative than Rabin in his use of sources. al-Sāmarrā'ī (1994) gives a scathing critique of these approaches, faulting in particular their reliance on ascription of variants to particular tribes, showing that depending on the source, the tribal ascription can vary significantly and this can render meaningless the entire project. I further develop this critique of tribes as a linguistic unit in Chapter 2.

More recently, it has become generally accepted within Arabic linguistics that Classical Arabic was based originally on a poetic koine (Versteegh, 2006), and therefore likely preserved a number of archaic features. Thus, it was probably distinct from the colloquial dialects even as it was beginning to be standardized.³ This situation is not unusual — around the world, poetic koinés tend to preserve archaic structures and vocabulary, as in Homeric Greek (Horrocks, 1997), the language of South Slavic epics (Foley, 1996), or the archaic Sanskrit of the Rīg Veda. These archaic structures would have been preserved by the formulaic and metrical

² Indeed, since the grammatical tradition tried to regularize and systematize the variation that they encountered, the later sources often give contradictory evidence versus the earlier sources. See the remarks in footnote 34 on page 300 regarding misinterpretations of an form which was not marked for case as representing a nominative part of a nominative/oblique form.

³ For example, the statement in Abboud-Haggar (*EALL: "Dialects: Genesis"*, p. 614): "Most scholars concerned with the issue of pre-Islamic Arabic agree that Arab tribes, whether Bedouin nomads dwelling in the Arabian desert or settled in sedentarized nuclei [...], spoke colloquial linguistic varieties, which to some extent differed from the variety used in poetry and the Qur'ān, and also from one another."

nature of Islamic poetry, as it has been in other languages with a diglossic poetic language.⁴

This shift in perspective means that there is no longer any reason to believe that a radical sociolinguistic change is responsible for the differences between Classical Arabic and modern colloquial varieties. Even more importantly, this change in perspective means that we can treat the modern Arabic colloquial dialects like we would any other language, with their own histories which are, to a large degree, independent of Classical Arabic and which we can therefore attempt to recover using the tools of historical linguistics.

0.1 Seeking Diversity in Arabic's Past

0.1.1 Previous Approaches

Though the studies detailed in the previous section are, strictly speaking, about the history of Arabic and of the diversity of Arabic, very few of them reconstruct earlier linguistic states of Arabic. Indeed, only a small handful of studies have actually attempted to do so. This section will review some of these studies, and discusses what elements of those studies were and were not helpful in developing the approach used here.

The earliest comprehensive study is that of Garbell (1958), who attempts to reconstruct the phonological development of 'Eastern Arabic' (essentially Levantine Arabic). She relies fairly heavily on written sources in Arabic and Aramaic, including sources with meta-linguistic commentaries, rather than on comparative reconstruction per se. One laudable element of this study is her focus on

⁴ For an overview of formulaic language in pre-Islamic poetry, see Monroe (1972) and for a modern example of metrical poetry in Arabic preserving forms foreign to a dialect, see Henkin (2010, chapter 5).

the chronology of those changes: She attempts to delineate five stages of changes from the pre-Islamic era to the twentieth century. However, her chronology is skewed by the assumption that Classical Arabic (or something very close to it) was the direct ancestor of modern Levantine Arabic colloquial dialects, and so the chronology of her account needs to be seriously reconsidered.

A more recent study is that of Owens (2006), but it is hard to discuss in terms of methodology since it is a collection of previously published articles with a variety of different approaches. The chapter closest in goals and aims to this dissertation is his Chapter 5 (originally published as Owens, 2005), which uses a large sample of modern dialects to investigate the diversity of what he terms ‘Pre-Diasporic Arabic.’ Owens is concerned here with the quantifiable amount of diversity present in pre-Islamic Arabic varieties, and uses a variety of computational techniques to calculate this relative diversity. Owens’ large sample is impressive, and is part of the inspiration for the sample here. However, this method does not reconstruct the forms that were present in ‘Pre-Diasporic Arabic,’ and it is not clear exactly how meaningful Owen’s mathematical dialectometry is for understanding the history of a language. Computer phylogenetic methods are somewhat better established (see Nichols and Warnow, 2008), but the results are difficult to interpret without established qualitative work in historical linguistics.

Cowan (1960) and Al-Jallad (2012) are both fairly traditional historical reconstructions of Arabic. Both focus primarily on reconstructing the ancestors of colloquial Arabic dialects. Cowan (1960) uses a sample of dialects from across the Arab world to perform a phonological reconstruction of what he terms ‘Proto-Colloquial Arabic’ and ‘Proto-Western Arabic’ using a word list reconstruction approach. It is impressive that he was able assemble comparative word lists for all the dialects in his sample. Dialect grammars rarely contain very much information

on vocabulary, let alone extensive enough word lists for comparative purposes. As a result of this, however, the size of the sample is basically limited to those dialects for which he could establish word lists, which means it is too small to capture very much of the diversity present in modern Arabic dialects. He also he did not have access to any of more recent sources that could shore up the data (e.g. the dialect atlases of Behnstedt, 1985, 1997; Behnstedt and Woidich, 1985).

Al-Jallad (2012) attempts to reconstruct the phonology and morphology of ‘Ancient Levantine Arabic’, what he regards as the ancestor of most modern Arabic dialects, contrasted apparently with the ancestor of other (undefined) dialects, primarily Najdi Arabic. His data is largely derived from pre- and early Islamic written texts (depending on how some of these texts are dated) and a small sample of Levantine dialects, complemented with data from Behnstedt (1997). His reconstructive approach to phonology does not appear to rely on comparative word-lists, but rather on individual lexical items found in his texts and grammars. He employs the Najdi dialect as a control, to represent a dialect that is supposedly not derived from this ‘Ancient Levantine Arabic’ ancestor. This is a problematic choice methodologically, since our documentation of Najdi Arabic is quite poor, and it is not necessarily clear how it is, or is not, related to other Arabic dialects. Indeed, though Al-Jallad does attempt to include examples and data from other dialects, the restricted focus of his approach means it is very difficult to understand what role exactly ‘Ancient Levantine Arabic’ played in the pre-history of Arabic.

Both of these approaches attempt to reconstruct a proto-language, and as I argue in Chapter 1, proto-language reconstructions are not the best tool for reconstructing the diversity that existed at a particular period. The ‘Ancient Levantine Arabic’ reconstructed by Al-Jallad seems to be picking up linguistic signals that are concentrated between about the fourth (the date of his earliest Arabic in-

scriptions) and thirteenth centuries CE (when Cyprus is said to ‘break off’ from Levantine Arabic), though of course various retained features would go back even farther. His choices of what data to include at times ignores different layers of the linguistic history of the Levant, so that, for example, he essentially ignores the demonstrative pronoun $\tau\gamma$ from the Namāra inscription in his reconstruction of the ALA demonstrative system (see further Section 6.1).

Similarly, Cowan’s (1960) reconstruction assumes the existence of a Proto-Western Arabic, which as we show in Section 4.2 is historically unlikely. Moreover, even though there is a certain homogeneity in North African Arabic dialects (including with respect to demonstratives, see Section 6.4), such a reconstruction itself doesn’t capture the diverse sources which must have contributed to those dialects, or how exactly they relate to the pre-Islamic diversity.

0.1.2 Current Approach

This dissertation diverges from these previous approaches, seeking to situate reconstructed diversity chronologically and geographically while using an approach which focuses on language as a tool of interaction between groups of people. The fundamental question which this dissertation seeks to answer is: “How is the diversity of modern Arabic dialects related to the diversity of Arabic dialects in the past?” Specifically, it sets out to reconstruct the dialect diversity which was present in Arabic immediately prior to the Islamic conquest, that is, the linguistic antecedents of modern Arabic dialects now spread across much of the eastern hemisphere. This is a massive undertaking, and so the more limited goal of the dissertation is to establish the groundwork that such a reconstruction can build upon, and to provide a proof-of-concept by reconstructing the diversity of pre-Islamic Arabic dialects in terms of their demonstrative system.

Similar to the approaches used by Cowan (1960) and Owens (2006), this dissertation reconstructs linguistic forms, specifically demonstratives, using a large sample of modern dialects. This contrasts with the approach taken by Al-Jallad (2012), and indeed this dissertation relies very little on data from Islamic era written sources, though I make some use of pre-Islamic epigraphic inscriptions.

One reason I chosen not to rely very much on written sources is the relatively early standardization of written Arabic, and the paucity of uncorrected or unrevised texts. This standardization appears to have been quite early, possibly before the earliest papyrus attestations available to us.⁵ Though early attestations of a language make up the bread and butter of the growing field of sociohistorical linguistics, the data simply is not nearly as rich for Arabic as it is for languages such as English (see e.g. Hernández-Campoy and Conde-Silvestre, 2012; especially the entry by Toulmin, 2012). Indeed, this early standardization means that the written language changes very little throughout its history and the relationship between the written Arabic language and the spoken, daily life registers is controversial and complex.

There is a fairly extensive literature on what is called ‘Middle Arabic’ an unfortunate term which is not chronological but rather refers to written Arabic which does not show all the features of Classical Arabic, though it is increasingly recognized that it is not tenable to view this register as a failed attempt at writing Classical Arabic (for an overview, see Lentin, *EALL : “Middle Arabic”*; see also the

⁵ For example, Hopkins (1984) discusses deviations from Classical Arabic in early Arabic papyri, which itself is a clear anachronism since Classical Arabic was largely not standardized at this point, but the relative homogeneity of these texts in comparison with, for example, many early written European languages is remarkable. Of course, Arabic certainly was being used for writing, probably on perishable materials, pre-Islamically, and so it is not surprising that enregisterment of the the written register would have begun prior to the establishment of Islamic states (see Macdonald, 2010, pp. 19-22).

very recent volume, E. Zack and Schippers, 2012, for the cutting edge in the field). Middle Arabic, while it might provide some clues about colloquial Arabic usage in the pre-modern era, is still a contentious area and so the results are not easy to incorporate into a historical study at this point. Only fairly recent texts, such as those from Egypt in the seventeenth century and later, tell us much straightforward information about the dialects, and this simply isn't enough time depth to be useful for the current study (see e.g. Davies, 1981; E. W. A. Zack, 2009).

0.1.2.1 Organization

The organization of this dissertation is as follows: This introduction lays out the justification and background for some of the decisions taken in designing this project, and presents a detailed breakdown of the data sources used. Chapter 1 establishes a theoretical framework which allows us to handle the fluid nature of human groups and linguistic diversity, and illustrates how and why this approach deviates from standard approaches in historical linguistics. Following this, Chapter 2 explores the social fabric of the pre-Islamic Arabian Peninsula and environs, particularly tribal organization and how this relates to modes of subsistence, in order to understand how Arabic-speaking groups interacted and spread. Chapters 3 and 4 provides the necessary and extensive historical background needed to understand which groups migrated where during and after the Islamic conquests. Chapter 5 then reconstructs the internal, linguistic history of the demonstratives in Arabic. Once all of these elements have been established, Chapter 6 combines them to reconstruct the speech communities which existed in the pre-Islamic and early Islamic Arabic-speaking world, and to show how those communities are related to the communities established after the Islamic conquests, and to the register of Classical Arabic. The conclusion reflects on the implications of this study for

major questions in Arabic linguistics, and proposes areas for further research.

0.2 Demonstratives

Demonstratives were chosen as the variable to explore present and past diversity in Arabic for a number of reasons. It is, of course, well established that morphology is a better indicator of genetic relationships than phonology (see the discussion in Hetzron, 1976), though we could have chosen a variety of other morphological features, for example, verbal conjugation or the personal pronouns. Beyond this, there are a number of advantages to using demonstratives as a variable that are specific to the case of Arabic.

First, Arabic demonstratives exhibit an incredible diversity. In Behnstedt (1985), a dialect atlas of what was previously Yemen Arab Republic (North Yemen), the authors distinguish some 54 different distal plural demonstrative forms (i.e. meaning ‘those’) in a country the size of South Dakota, with the population density of Alabama (about 95 per square mile). In Upper Egypt, villages separated by a few kilometers have wildly divergent distal demonstrative forms, but as far as the data shows, almost identical proximal demonstratives. In North African dialects, in contrast, the demonstratives are remarkably uniform in spite of multiple population movements into the area over thirteen hundred years.

Second, since there are still no sufficiently detailed historical phonological studies of Arabic, it is difficult to focus on forms whose variation seems to be at heart phonological. Verbal conjugation, for example, tends to be quite similar between Arabic dialects, and many of the differences which do exist are a result of either sound changes or of synchronic differences in phonotactics. With demon-

stratives, even without knowing the underlying sound changes involved, it is generally straightforward to determine what the original forms of the demonstratives were. This means, of course, that we cannot compare certain aspects of the demonstratives — for example, the distal suffix *-k* sometimes has a final */-a/* sound, but not always, and this is probably as much a question of the historical phonology of a given dialect as of developments within the demonstratives. Until we have a better sense of the sound changes that occurred in Arabic, we will have to be content with relatively broad comparisons during morphological reconstruction.

Third, demonstratives in Arabic are paradigmatic, marking gender, number and distance contrasts, which means we know to expect certain types of leveling and analogical processes. Moreover, the paradigms also seem to record the results of borrowing and inter-dialect shift, since higher frequency demonstrative forms (either singular or proximal forms) tend to be borrowed more quickly (as they are more available in the linguistic input) than lower frequency forms. This means that demonstrative paradigms give us a clear picture both of the origins of a dialect, and its more recent development.

Finally, the demonstratives in Arabic represent something of a *terra incognita*. Sources on Arabic dialects frequently offer ‘reconstructions’ of the demonstratives, but these are simply decompositions of the demonstratives into (apparent) pseudo-morphological components, and presented as if they are some sort of compounds, in spite of the fact that compounding processes are vanishingly rare in Arabic. A good example of this comes from the standard reference *Handbuch der arabischen Dialekte* (W. Fischer and Jastrow, 1980, p. 82):

Mit dem vom Verbum stammenden Pluralsuffix *-ū* entstanden die Formen *ḏū*, *hāḏū*, **ḏā-ū* > *ḏāw*, *hā-ū-ḏā* > *hawḏā*. [...] Das aar. *ʿulā* steckt

in den Verbindungen $hā-ʿulā > hawlā$ $*(hā)ḏā-ʿulā > (hā)ḏawla$; eine Variante $*ʿilā$ liegt in $*ḏā-ʿilā > ḏaylā$ vor.

With the plural suffix *u:*, coming from the verbs, the forms $ḏu:$, $ha:ḏu$, $*ḏa:-u: > ḏa:w$, $ha:-u:-ḏa: > hawḏa:$ are derived. The Old Arabic $?ula:$ is placed in the compound $ha:-?ula: > hawla:$ $*(ha:)ḏa:-?ula: > (ha:)ḏawla$; A variant $?ila:$ is attested in $*ḏa:-?ila: > ḏayla$.

Such a treatment raises more questions than it answers: How did a verbal ending end up in the middle of a demonstrative, a nominal form? Are these elements *ha:* and *ḏa:* equivalent? If not, how is it that they undergo the same processes of affixation/compounding? What are these two variant $?ula:$ and $?ila:$ forms, why is only one listed as belonging to Old Arabic, and what is their relationship to one another?

These treatments can be confusing, since they are often given without further explanation. For example, Al-Jallad (2012, p. 324, fn. 372) cites Jastrow (1978, p. 105) as claiming that modern *hawdi* ‘these’ is derived from $*ha:-u:-ḏi:$ but notes, “[i]t is unclear to me what the source of the *u:* formative is.”

On a more fundamental level, however, this type of account also seems incongruous given what we know about Arabic and other Semitic languages. These languages do not generally use compounding as a productive strategy, and suffixes and prefixes are often relatively transparently derived from originally independent words. Thus, one of the primary goals in this dissertation’s analysis of the demonstratives in Chapter 5 is to explain how, in a language where compounding processes are rare, demonstratives came to be composed of strings of morphemes.

0.3 Data Sources

The procedure for understanding the data in the study rests on a multi-phase approach. The primary sample, composed of dialect grammars and dialect atlases, and which comprises sixty-three sample points, is used to compose a database, shown in the appendix. The database is used to generate initial hypotheses about how a feature developed over time and the implications for the grouping of dialects. Then, dialect atlases are used to understand the dialect geography of the demonstrative forms. These atlases also provide a greater number of sample points which can show possible counterexamples to the groupings, or can provide more information on the sampled points from the first phase, e.g. whether they represent transitional zones between dialects where there is a blending of different features, etc. Finally, the secondary literature is consulted to ensure that the findings are in line with previous works.

The sampling method here differs from many previous studies, such as those of Al-Jallad (2009) and Owens (2006) in that it does not attempt to privilege any particular sample points, or bias the sample towards a given type of dialect due to assumptions about whether those dialects are more or less archaic than others. If we wish to understand the development of Arabic, all dialects should be taken into equal consideration — they will represent different layers of development, and the researcher’s job is to tease apart those layers as far as is possible.

The sample was designed to offer geographically balanced coverage of the Arabic speaking world, and to include dialects which are classified as belonging to different ‘types’ according to current classification schemes (e.g. ‘Bedouin’ versus ‘Sedentary’ or ‘Pre-Hilalian’ vs. ‘Hilalian’ dialects). There are limits to how much coverage is possible — for example, Iraq is presumably one of the most important areas for understanding the history of Arabic dialects, as it was settled

pre-Islamically, and was the center of the Arabo-Islamic world for over half a millennium. However, it has been politically unstable for much of the twentieth century, and so the dialectology of Iraq has been poorly recorded. What sources are available on Iraqi dialects are often published in piece-meal or difficult to access sources, and so I discovered many of these sources only after the sample had been assembled.⁶

This sampling approach is especially important for working with paradigmatic morphology, since different morphological subsystems influence one another. For example, since demonstratives are often heavily influenced by pronouns, a sampling approach makes it easier to quickly check whether this influence has occurred in a given dialect, even if at this stage we are not reconstructing the pronominal system. As we reconstruct more of the morphology of the dialects, having a stable sample makes it easier to compare results from one area with those from another. For this reason, whenever possible I have tried to take data from general descriptive grammars for each sample point. Articles from the *Encyclopedia of Arabic Language and Linguistics* (Versteegh, 2005 – 2009) were especially useful since they are easy to access and have a unified structure.

I have also tried as much as possible to rely on grammars of small groups, preferably villages, rather than regional or national areas. This way I hope to ensure that the samples are of actual paradigms as they would exist in the ideolects of most of the speakers in that area or group. Dialect grammars of large areas often give several variant forms for each demonstrative, but it is rarely clear if these

⁶ The documentation of Arabic dialects is even worse for the Arabian Peninsula generally, and especially for areas now part of the Kingdom of Saudi Arabia. One of the major goals for Arabic dialectology in the twenty first century should be to better document Saudi Arabian and other Peninsular dialects. With increasing numbers of Saudi researchers working on Arabic linguistics, I hope that we will soon have access to ever better information on Peninsular dialects.

different variants represent regional, sociolinguistic or pragmatically determined variation and which forms are part of the same paradigms.

This sampling approach contrasts with the sampling approach of the only other major work on Arabic demonstratives, W. Fischer (1959). In that work, he simply gathers any and all reports of demonstratives in the literature, but not necessarily data about other subsystems of these dialects. Nor does he try to ensure that these forms represent intact paradigms. That said, W. Fischer (*ibid.*) has been the primary secondary source used to supplement the data from the sample here, though there have been many developments in Arabic dialectology since his time. His primary goal is to produce a typology of demonstrative forms based on structural and functional similarities, though he does propose some historical reconstructions. The advantages of his approach is that it has significantly better coverage than a restricted sample (though there are now many more available sources), but it does not give as clear a view of the paradigmatic pressures on these demonstratives. I have used Fischer's monograph primarily as a way to expand the scope of the data, so that it is easier to know whether or not a given structure is attested in Arabic, as well as to get further insight into difficult to reconstruct constructions.

Fischer's different approach also produced somewhat different results. His categorization of dialects is primarily based on whether or not they have developed **ha:-*prefixed demonstratives (*ibid.*, pp. 36-7). In contrast, I argue that the historical development of Arabic demonstratives is primarily reflected in the latter part of the demonstratives, what I call the **ḍa:-*part, and that the attachment of the **ha:-*prefix to the demonstratives typically occurred later, sometimes through contact, such that the **ha:-*prefix actually spread from dialects with that prefix to dialects that lacked that prefix (i.e. 'areal spread'). Nonetheless, Fischer's data, as well as

his analyses, have been invaluable for this study.

Other sources are also used when relevant. Inscriptions in Old Arabic or other Semitic languages are used to understand both the development of the demonstratives and the development of different speech communities. Al-Jallad (2012) has been an invaluable source for these inscriptions and epigraphic materials, since he collects most of the materials which Semiticists have treated as Arabic, though I do not always agree with his language-essentialist approach to classifying them, as explained in Section 3.3.1. For a broader Semitic overview of the developments of the demonstratives, I have relied on Hasselbach (2007a), which is the most recent treatment of this subject.

0.4 Sample points

0.4.1 Classical Arabic

Classical Arabic is not a linguistic entity whose systems are matched, i.e. they do not reflect the evolution of a single variety. Rather, it is a composite variety, a koine in the original sense of the word, a constructed variety which reflects a number of inputs from different varieties which were not necessarily included in a systematic matter. Moreover, the various systematizations of Classical Arabic, as reflected in reference grammars (both ancient and modern) and in Modern Standard Arabic, tend to hide much of the variety that was present even in the pre-Islamic Koine.

At the same time, Classical Arabic may be viewed as an eclectic collection of features which were present in various and variously related, pre-Islamic dialects, a sort of archive of ancient forms. Thus, it can cautiously be used to date the existence of various forms that were present in the language in the pre-Islamic era. This time depth estimation however only gives us a cut-off in a sense - we do not

know *how* ancient these forms are, and many (for example, dual verb conjugations) seem to have been largely absent from the dialects that formed the basis of the Islamic expansion, suggesting that these forms were already quite archaic at that time of the Islamic expansions. Alternately, some of the dual forms (e.g. dual verbs) could have been innovated by analogy with other dual forms (e.g. nominals) and these innovations could be specific to Classical Arabic.

Indeed, since the focus here is on the direct ancestors of the modern Arabic dialects, features which are found only in Classical Arabic, but not in the modern dialects, will not be reconstructed for the ancestors of those dialects unless there is good evidence for the existence of those forms. The best example of this is the dual demonstrative forms which are said to be a part of Classical Arabic, but which are not attested in any modern dialects. Though there are forms which appear similar, these are often just surface similarities. For example, one could conceivably argue that a form like *ḍanna* (this.F.PL.) derives from the dual, i.e. *ḍa:na* (this.DUAL.M.NOM) > *ḍanna* (w)ith the moraic structure preserved by doubling the consonant (cf. demonstratives like *hayy* < **ha:y*). Nonetheless, this form actually appears to be derived from *ḍalla* based on the distribution of that form, with a somewhat but not entirely unusual /l/ > /n/ development. We would also have to account for the semantic change from a masculine dual form to a feminine plural form (see further Section 5.3.2.2). Therefore, for purposes of this study, the existence of a form in Classical Arabic is not sufficient evidence by itself to reconstruct it back to the ancestor of a given dialect.

0.4.2 Peripheral Dialects

There are a small number of peripheral dialects in Arabic which have often received an outsized amount of attention from Arabic linguists, due to their status

as dialects which have been ‘cut off’ from the sociolinguistic context of the mainstream of the Arab world. In particular, these dialects are often cut off from the Arabic-language literary and cultural tradition, meaning they don’t participate in the same traditions of diglossia, though of course as separate speech communities they are also helpful in dating when certain changes occurred in their parent dialects. However, the special status accorded to these dialects in some studies (e.g. Al-Jallad, 2009; Owens, 2006) is exaggerated, as we often don’t know how cut off these dialects were exactly (human migration and interaction tends to be underestimated in general), and the cutoff dates are rarely used to any great effect in systematically dating changes.

I have grouped these dialects together here to reflect the tradition of modern Arabic linguistics, though in general these dialects pattern quite closely with their most geographically proximate neighbors.

The dialect of the Maronites of Cyprus (CYP) is represented here by:

Alexander Borg (1985). *Cypriot Arabic*. Stuttgart: Kommissionsverlag Franz Steiner Wiesbaden GMBH

Alexander Borg (2006). “Cypriot Maronite Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 536–543

The history of Malta and its Arabization is discussed in 4.4.1, though again it is worth noting that in spite of the reconquest of the island in 1091 CE, Classical Arabic continued to be used even in inscriptions (on e.g. headstones) until the twelfth century, and the ‘cut-off’ between Malta and Arabic-speaking areas is probably better dated to the thirteenth century when Muslims were expelled from Sicily and Malta. The dialect of Malta (MALT) is represented by:

Peter Schabert (1976). *Laut- und Formenlehre des Maltesischen anhand zweier Mundarten*. Verlag Palm & Enke Erlangen

Manwel Mifsud (2006). “Maltese”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 146–159

The Anatolian dialects of Arabic are said to have been cut off from the mainstream of Arabic dialects also (see e.g. Borg, *EALL : “Cypriot Maronite Arabic”*, p. 537), though it is not clear when this happened, and given the lack of specific barriers, unclear how strict separation could be maintained. These dialects are represented here by the dialect of Daragözü (ANDAR), based on:

Otto Jastrow (1973). *Daragözü: eine arabische Mundart der Kozluk-Sason-Gruppe (Südostanatolien)*. *Grammatik und Texte*. Nürnberg: Verlag Hans Carl

Supplemental data on Anatolian dialects is taken from:

Otto Jastrow (1978). *Die Mesopotamisch-Arabischen Qeltu-dialekte. Band I: Phonology und Morphologie*. Wiesbaden: Kommissionsverlag Franz Steiner GMBH

The Chadian, Nigerian and Sudanese dialects will be discussed below, since they are clearly historically derived from Nile Valley dialects. The dialects of Uzbekistan (UZBA and UZBDJ) are drawn from:

Gerit Zimmermann (2006). “Uzbekistan Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 612–623

The dialect of Afghanistan (AFGA) is drawn from:

Bruce Ingham (2006). “Afghanistan Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 28–35

0.4.3 North African Dialects

Though the distinction between Hilalian and Pre-Hilalian dialects has been criticized in Chapter 4.2.1, I report the categorization chosen by the authors of the individual grammars below. Dialects are given from east to west.

The dialect of Benghazi and eastern Libya (NABEN) is represented by:

Jonathan Owens (1984). *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrasowitz

He makes no comment on whether the dialect is pre- or post “Hilalian”, though Pereira (*EALL* : “*Libya*”) states that “as a whole, the Libyan Arabic dialects are of the Bedouin type” and opposes this to Hilalian dialects. Supplemental information for Libyan Arabic is taken from:

Christophe Pereira (2006). “Libya”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 52–58

The Jewish dialect (NATRJ) of Tripoli is represented by:

Sumikazu Yoda (2005). *The Arabic Dialect of the Jews of Tripoli (Libya)*. Wiesbaden: Harrasowitz Verlag

The Muslim dialect (NATrM) is represented by:

Christophe Pereira (2006). “Tripoli Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 548–556

The latter source sees the dialect of Tripoli as representing a pre-Hilalian dialect with significant Hilalian influence, and the former asserts (p. 1) that the Jewish dialect is likely more resistant to later changes than the Muslim dialect.

Only the dialect of Jews in Tunis (NATUN) is included here, based on:

Hans-Rudolf Singer (1984). *Grammatik der arabischen Mundart der Medina von Tunis*. Berlin: Walter de Gruyter

The modern dialect of the Muslims of Tunis is seen as a modern koine showing traits from the rapid urbanization of the city (Gibson, *EALL* : “*Tunis Arabic*”), and so the well described Jewish dialect, seen by Singer as a pre-Hilalian dialect, seems a better choice to use here. Some comparative reference is made to:

Maik Gibson (2006). “Tunis Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 563–571

David Cohen (1970). “Les deux parlers arabes de Tunis”. In: *Études de linguistique sémitique et arabe*. The Hague and Paris: Mouton, 150–171

The pre-Hilalian dialect of Djidjelli (NADj) is represented by:

Philippe Marçais (1956). *Le parler arabe de Djidjelli, Nord constantinois, Algérie*. Paris: Librairie d'Amérique et d'Orient

The dialect of the Mzāb Algerian desert (NAMz) is represented by:

Jacques Grand'Henry (1976). *Les Parlers Arabes de la Région du Mzāb (Sahara Algérien)*. Leiden: E.J. Brill

This region has served as a refuge for Ibaadite refugees beginning in the eleventh century CE, after their village in the mountains between Oran and Algiers (near modern Tiaret) was sacked by the Fatimids in 909 CE. In 1075, the Ibadites founded five villages in the Mzāb region, though nomadic tribes in the region claim a link to the Hilalian migrations and the study includes members of both sedentary and nomadic groups (*ibid.*, pp. 2, 6).

The pre-Hilalian dialect of Cherchell (NACHR) is represented by:

Jacques. Grand'Henry (1972). *Le parler arabe de Cherchell, Algérie*. Louvain-la-Neuve: Université catholique de Louvain, Institut orientaliste

The pre-Hilalian dialect of Tlemcen (NATL) is represented by:

William Marçais (1902). *Le Dialecte Arabe Parlé à Tlemcen: Grammaire, Textes et Glossaire*. Paris: Publications de L'École des Lettres D'Alger

The dialect of Fez (NAFÈZ) is drawn from:

Dominique Caubet (1993). *L'Arabe Marocain: Phonologie et Morphosyntaxe*. Paris-Louvain: Éditions Peeters

It is not clear whether this should be considered a pre-Hilalian or post-Hilalian dialect. Note that Heath (2002) finds that Moroccan dialects are very homogeneous when it comes to the demonstratives, most showing the same forms as those of Fez.

The dialect of Mauritania (NAHAS), called Ḥassāniyya, is considered a Hilalian dialect. Two sources were used for this dialect:

David Cohen and Mohammed el Chennafi (1963). *Le dialecte arabe ḥassānīya de Mauritanie, parler de la Gabla*. Paris: C. Klincksieck

Catherine Taine-Cheikh (2006). “Ḥassāniyya Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 687–699

0.4.4 Nile Valley and Southern African Dialects

Egypt generally is represented by the dialect maps in:

Peter Behnstedt and Manfred Woidich (1985). *Die ägyptisch-arabischen Dialekte*. In 5 volumes. Weisbaden: L. Reichert

However, unlike Yemen and Syria, the dialect maps are not as amenable to the system of selecting a series of sample points in Egypt, and therefore the maps were used in a non-systematic manner to explore the dialect geography of Egypt.

The dialect of Cairo (EGCAI) is represented by:

Manfred Woidich (2006b). *Das Kairenisch-Arabische. Eine Grammatik*. Wiesbaden: Harrassowitz Verlag

The Upper (southern) Egyptian dialect of the il-Bi‘erāt region, i.e. B‘eri Arabic, (EGBr), on the West Bank of the Nile near Luxor, is represented by:

Manfred Woidich (2006). “B‘ēri Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 299–308

The eastern Sudanese dialect of Shukriyya (SHUK) is represented by:

Stefan Reichmuth (1983). *Der arabische Dialekt der Šukriyya im Ostsudan*. New York: Georg Olms Verlag

Nigerian Arabic (NIG) is represented by:

Jonathan Owens (1993). *A Grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz Verlag

Mainstream Chadian Arabic (CHD) is represented by:

Claude Hagège (1973). *Profil d’un parler arabe du Tchad*. Paris: P. Geuthner

The dialect specifically of the Shuwa (CHDSH) tribe is represented by:

Alan S. Kaye (1976). *Chadian and Sudanese Arabic in the Light of Comparative Arabic Dialectology*. The Hague: Mouton

Some additional information is provided by:

Patrice Jullien de Pommerol (2006). “Chad Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 360–368

0.4.5 Levantine Dialects

The Lebanese dialect of Bishmizzīn (LvBSH) is represented by:

Michel Jiha (1964). *Der Arabische Dialekt von Bishmizzin*. Wiesbaden: Franz Steiner Verlag

The dialect of Damascus (LvDAM) is represented by:

Jérôme Lentin (2006). “Damascus Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 546–555

Mark W. Cowell (1964). *A Reference Grammar of Syrian Arabic, based on the dialect of Damascus*. Washington, DC: Georgetown University Press

The dialect of Hebron (LvHEB) is represented by

Ulrich Seeger (1996). *Der arabische Dialekt von il-Xalil (Hebron)*. Possibly unpublished manuscript. Karlsruhe

Many of the dialects in Syria were chosen by collating the dialects maps in:

Peter Behnstedt (1997). *Sprachatlas von Syrien*. Weisbaden: Harrassowitz

As opposed to Yemen, it was too difficult to get a member of every single dialect group, but careful perusal of the dialect maps supplemented these sample points. The sample points were, approximately from north to south, with the dialect group they represent according to Behnstedt’s classification in parentheses:

- LvS101: Milles (Group B)
- LvS239: ʿAyn əš-Šariye (Group C)
- LvS281: Xanāšir (Group IIIb)
- LvS305: Kfar Bō (Group G)
- LvS330: Šanšar (Group H)
- LvS340: Ḥuwwārīn (Group L)
- LvS348: Ras əl-Maʿarra (Group T)
- LvS417: Saqqa (Group Z)
- LvS433: Ḥamer (Group Y)

The dialect of Palmyra (LvSPAL, Behnstedt's group K) was drawn both from the dialect maps and from:

Jean Cantineau (1934). *Le Dialecte Arabe De Palmyre*. Vol. 1. Beirut: Mémoires de L'Institut Français de Damas

The dialect of Soukhne (LvSSou) was drawn from dialect maps and from:

Peter Behnstedt (1994). *Der arabische Dialekt von Soukhne (Syrien)*. Semitica viva. Harassowitz

0.4.6 Mesopotamian Dialects

The dialects of Iraq are sadly poorly documented, largely due to the turbulent history of that country in the 20th century. What dialectological work has been done has largely been conducted outside that country, particularly among refugees from various religious minority groups.

The dialects of Baghdad, both the Muslim *ḡalāt* (IQBGG) and the Jewish *qəltu* dialect (IQBGJ) were drawn from:

Haim Blanc (1964). *Communal dialects in Baghdad*. Cambridge, MA: Harvard University Press

The dialect of the Jews of ʿAqra (IQAQ) in Northeast Iraq was represented by:

Otto Jastrow (1990). *Der arabische Dialekt der Juden von Aqra und Arbil*. Weisbaden: O. Harrassowitz

The dialect of the Jews of Mosul (IQMosJ) was represented by:

Otto Jastrow (1979). “Zur Arabischen Mundart Von Mossul”. In: *Zeitschrift für Arabische Linguistik* 2, pp. 36–75

The dialect of Iraqi Khuzestan (IQKHZ) was represented by:

Bruce Ingham (2006). “Khuzestan Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 2. Leiden: Brill, pp. 571–578

It was supplemented with data from:

Bruce Ingham (1973). “Urban and Rural Arabic in Khūzistān”. English.
In: *Bulletin of the School of Oriental and African Studies, University of London* 36.3, pp. 533–553

Supplemental information on Iraqi dialects was drawn from:

- Otto Jastrow (2006). “Iraq”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 2. Leiden: Brill, pp. 414–424
- T. M. Johnstone (1975). In: *Annals of Leeds University Oriental Society* 3, pp. 89–109
- Erkki Salonen (1980). “On the Arabic Dialect Spoken in Šīrqāṭ (Assur)”. In: *Suomalaisen tiedeakatemian toimituksia annales academiae scientiarum fennicae* 212
- G. Khan (1997). “The Arab Dialect of the Karaite Jews of Hīt”. In: *Zeitschrift für arabische Linguistik* 34, pp. 53–102
- Qasim R. Mahdi (1985). “The spoken Arabic of Baṣra, Iraq”. Unpublished Ph.D. thesis. The University of Exeter

0.4.7 Arabian Dialects

Two sample points were chosen from the dialects of the Saudi Arabian Tihama: the dialects of Qahabah (AQQAḤ) and Al-Qauz (ARQZ). Data for these dialects was taken from:

Theodore Procházka Jr. (1988). “Gleanings from southwestern Saudi Arabia”. In: *Zeitschrift für Arabische Linguistik* 19.1, pp. 44–49

The dialect of Oman (AROM) is represented by:

Lutz Edzard (2006). “Omani Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 478–491

The dialects of Bahrain, both the Baḥārna dialect (ARBAH) and the ‘Anzī dialect (ARANZ) are represented by:

Mahdi Abdalla Al-Tajir (1982). *Language and Linguistic Origins in Bahrain*. London: Kegan Paul International

0.4.8 Yemeni Dialects

The following eight data points were chosen in Yemen:

- Y104: im-‘Āgir (Region 8: South East Dialects)
- Y121: Kuhlān (Region 7b: Southern Plateau)
- Y145: Ta‘izz (Region 2a: k-dialects)
- Y156: Jabal am-Nār (Region 1c: Tihāma Dialects)
- Y24: Ḥāmideh (Region 5a: Northern Plateau)
- Y6: an-Naḍīr / Jabal Rāziḥ (Region 2: k-dialects)
- Y96: Bayt al-Faqīh (Region 1b: Tihāma Dialects)
- Y99: Gā’ideh (Region 2b: k-dialects)

Their paradigms were recorded by collating the maps in:

Peter Behnstedt (1985). *Die nordjemenitischen Dialekte*. Wiesbaden: L. Reichert

Additionally, three villages in northern Yemen were also included in the sample, since the dialect of im-Talḥ (YTALH), the dialect of Wasīt (YWas) and the dialect of Xašir (YXASH). These dialects were included since the following source gives examples of usage:

Peter Behnstedt (1987). *Die Dialekte der Gegend von Sa‘dah (Nord-Jemen)*. Wiesbaden: Otto Harrassowitz

Finally, the dialect of the Ḥadramawt area in south-eastern Yemen was represented by:

Abdullah Hassan Al-Saqqaf (2006). “Wādī Ḥaḍramawt Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 687–699

0.4.9 Summary Chart and Map of Sample Points

Table 0.1 gives a summary of the abbreviations for each sample point used in the dissertation. Figure 0.1 shows the sample points on a scalable vector map, so the reader can zoom into areas with dense sampling using their PDF viewer.

Table 0.1: Summary of Dialect Abbreviations

Abbreviation	Dialect Name
AFGA	Afghani
AND	Andalusi
ANDAR	Daragözü (Anatolia)

Table 0.1: Summary of Dialect Abbreviations (continued)

ARANZ	‘Anzī (Bahrain)
ARBAH	Baḥārna (Bahrain)
AREM	Emirati
ARNJD	Najdi
AROM	Omani
ARQAH	Qahabah
ARQZ	Al-Qauz
CHD	Chadian (general)
CHDSH	Chadian (Shuwa)
CYP	Cypriot
EGBR	B‘ēri (Egypt)
EGCAI	Cairene
IQAQ	Jewish ‘Aqra
IQBG	Baghdadi <i>gəlat</i> (Muslim)
IQBJ	Baghdadi Jewish
IQKHZ	Khuzestani
IQMOSJ	Jewish Mosul
LvBSH	Bishmizzīn (Lebanon)
LvDAM	Damascus (Syria)
LvHEB	Hebron (Palestine)
LvS101	Milles (Syria)
LvS239	‘Ayn əš-Šarīye (Syria)
LvS281	Xanāšir (Syria)
LvS305	Kfar Bō (Syria)
LvS330	Šanšar (Syria)
LvS340	Ḥuwwārīn (Syria)
LvS348	Ras əl-Ma‘arra (Syria)
LvS417	Saqqa (Syria)
LvS433	Ḥamer (Syria)
LvSPAL	Palmyra (Syria)
LvSSOU	Soukhne (Syria)
MALT	Maltese
NABEN	Benghazi (Libya)
NACHR	Cherchell (Algeria)
NADJ	Djidjelli (Algeria)

Table 0.1: Summary of Dialect Abbreviations (continued)

NAFeZ	Fez (Morocco)
NAHAS	Hassaniyya (Mauritania)
NAMz	Mzāb (Algeria)
NATL	Tlemcen (Algeria)
NATrJ	Jewish Tripoli (Libya)
NATrM	Muslim Tripoli (Libya)
NATUN	Jewish Tunis
NIG	Nigerian
SHUK	Shukriyya (Sudan)
SINR	Sinai Rmēliy
SINS	Sinai Smē'niy
UZBA	Uzbekistan Arabkhona
UZBDJ	Uzbekistan Djogari
Y104	im- ^ʿ Āgir
Y121	Kuhlān
Y145	Ta'izz
Y156	Jabal am-Nār
Y24	Ḥāmideh
Y6	an-Naḍir / Jabal Rāziḥ
Y96	Bayt al-Faqīh
Y99	Gā'ideh
YHADR	Hadramawt
YTALH	im-Talḥ (Yemen)
YWAS	Wasīt (Yemen)
YXASH	Xašir (Yemen)



Figure 0.1: Map of the sample dialects. Sizes differ only for purposes of clarity and do not signify relative size or importance. The reader can zoom into areas with smaller fonts in their PDF reader.

0.5 Conventions and Abbreviations

Italics are used for linguistic examples from living languages, or languages for which we know the vowels. Small caps are used for linguistic examples for which we only have the consonantal skeletons, a result of the use of abjads to write Semitic languages. Small caps with alternating upper-and-lower cases are used for the code names of Arabic dialects used in the sample here, and for the first use of an important term. In-text examples have italics for the forms in the language, followed optionally by a gloss in parenthesis, and a free translation in single quotes. Translation of foreign language material is in single quotes, which are also used for 'scare quotes'; double quotes are reserved for quotations in running text. All linguistic examples are transcribed in IPA, rather than the idiosyncratic systems used to transliterate Semitic languages, but proper names are written in their conventional form or using Semitic-style transcriptions, following the general convention of the *Zeitschrift für Arabische Linguistik*. I translate examples glossed in other sources but typically put the original in an accompanying footnote, though there are times when I will translate the Arabic directly.

Where dates are given in both the Islamic calendar and the Gregorian calendar, the Islamic date is given first, followed by the Gregorian date after a slash mark, e.g 12/633.

Table 0.2: List of conventions

<i>word</i>	Linguistic examples in languages with known vowels
KLM	Linguistic examples without known vowels
LvS101	Name of sample dialect
<i>kalima</i> (word.INDEF) ‘a word’	<i>Example</i> (Gloss) ‘Translation’
ALIGNMENT EVENT	First use of important term
‘dubious concept’	Quotes of terminology or concepts; translations; ‘scare quotes’
“It is well established that [...]”	Direct quotes
/i/	Specific phoneme
[i]	Specific phone or phonetic realization
Q, J	Phonemes with sociolinguistically variable pronunciations, i.e. variables

Chapter 1

From the genealogy of languages to the archeology of speech communities

The purpose of this chapter is to develop a model of linguistic reconstruction which can be used to discover not the pre-history of a language, but rather specific phases in a language's history in order to better understand how the modern diversity of that language is related to the diversity which existed at a particular time period. This framework is heavily inspired by that of Ross (1997, 2005), Toulmin (2009), Enfield (2003) and Croft (2000).

There are two major factors which must influence the design of this model: The first is the goal of the reconstruction, which is to reconstruct a particular historical period, and which therefore requires a somewhat different approach than methods that are designed to reconstruct more distant proto-languages. The second is the nature of the Arabic data. We have a massive amount of dialect diversity, but no stable, standardized entities that can be conveniently selected as inputs to reconstruction. We know that there was frequent and long-term contact between

these dialects throughout their histories. Moreover, the epigraphic data for Arabic is complex, remarkably homogeneous and sparse in the pre-Islamic era, but on the other hand there is a rich historical tradition which can be leveraged to understand the history and relations of different groups.

These factors — the need for a historically situated reconstruction, and the nature of our Arabic data — are challenging for the standard comparative method of reconstruction to handle, and the first section of this chapter explores these challenges. The following sections propose a different approach based on speech communities, which allows us to leverage historical information to situate reconstructed entities geographically, historically and socially. Building on the works of the authors cited above, who define speech communities primarily in terms of their network connections (see Ross, 2005, for a clear example of this), I use data from sociolinguistics to argue for another defining characteristic of speech communities, the allegiances which speakers have towards different communities.

The value of speech communities, I argue, is that they are what shape and guide the extent of how changes move within and across languages. Furthermore, since speech communities are coincident with communities more generally, they can be reconstructed using non-linguistic evidence from historical and literary sources, while anthropological and geographical literature can help adduce what kinds of social structures likely existed during the history of those communities. Moreover, the linguistic results of contact can help us understand the relations between the speech communities in contact, and thus relate this to our historical data.

1.1 Challenges in applying the comparative method

1.1.1 Units of analysis

One of the most fundamental questions in reconstructing linguistic change over time is what our unit of analysis must be. The unit of analysis is both the input to and the output from the process of reconstruction. Within linguistics generally, but particularly within historical linguistics, the unit of analysis is a ‘language’, a unit defined in theory at least by its linguistic components. There will be many manifestations of these linguistic components – text records (both modern and ancient), grammatical descriptions, actual speech in interaction – but they are all assumed to be manifestations of the abstract unit of language, which somehow subsumes all of the linguistic data that makes these manifestations possible. Even dialect variation might be subsumed within the larger unit of language, as we often hear that in dialect X of language Y (e.g. in the Iraqi dialect of Arabic), though of course the relations between these levels of analysis is one of the great ongoing controversies in linguistics.¹

Within historical linguistics, language is defined at least in theory as a variety or set of varieties which exclusively share certain common historical developments.² Thus, a language in this tradition is defined by the core changes which distinguish it from its immediate relatives, and subdivisions of that language are all united by these core sets of changes.

This definition of language does not fit all situations equally well. First of all, a proper reconstruction would be necessary before this definition could be

¹ Ross (1997) has proposed using the term *LECT* to avoid the problem of how to distinguish between ‘dialect’ and ‘language’, but this is still a unit defined primarily by linguistic considerations.

² See for example how Al-Jallad (2012, pp. 76-81) defines Arabic based on its descent from Semitic, a very reasonable way to define a language in this framework.

applied, and the researcher would be caught in a problem of the chicken and the egg.

Second, for languages which are mostly attested in written sources (i.e. dead languages, or those with few surviving modern descendants) it is easier to maintain this purely essentialist linguistic definition. Time, and the fact that writing is a process mediated by various levels of standardization and codification,³ will hide much of the variation that was present in earlier stages of the language. This maintains what may simply be an illusion of linguistic unity, but that illusion can be helpful if the researcher is less interested in modern diversity and more interested in reconstructing further and further back in time.⁴

However, as one moves closer to the present, the diversity that has been hidden by this approach becomes increasingly apparent, and the historical linguist, with her clear cut trees, morphs into the dialectologist, with endless maps of cross-cutting isoglosses. Often, the dialectologist is at a loss to come up with a means of choosing some synchronic unit based on her data that seems sufficiently focused to use as any kind of unit of analysis.⁵

Indeed, the notion of a relatively unified and homogeneous variety that can be called a ‘language’ simply does not seem to exist in a meaningful way in many communities, and so there is no clear place for the researcher to begin. A strong argument in this vein is that of Mühlhäusler (1996, p. 7) who, in studying

³ Orthography also can play a role in hiding diversity, as with the Semitic adjads where vowels are not written, and so we have little sense of how vowel qualities related to diversity in old Semitic languages.

⁴ H. Andersen (1996) refers to proto-languages as “convenient fictions” but we might say the same things about ‘languages’ more generally.

⁵ This problem is particularly evident in attempts to classify Arabic dialects, such as those in Palva (*EALL: “Dialects: Classification”*), where general trends for particular dialect types are about all that the author can identify, each of which is rife with exceptions.

the Pacific, concludes that:

[T]he notion of ‘a language’ is one whose applicability to the Pacific region, and in fact to most situations outside those found within modern European type nation-states, is extremely limited.

Similarly, LePage’s (1992) work on societies in the midst of political and linguistic transitions shows that the linguistic situation in a given area might not be sufficiently focused for a linguistic definition of ‘language’ to be meaningful:

Those of us who come from very stable and highly focused societies may find it difficult to distinguish stereotypes about “normal transmission” from the real facts about language use, variation, and change in use, since we are so accustomed to think in terms of idealized, reified, discrete systems; it is essential to see all language questions in terms of activity between individuals as they form social groups, even in the most static and highly focused societies. (p. 98)

In reality, historical linguists often do not actually stick to a truly linguistically defined unit of analysis. Instead, the choice of what to analyze is made on the basis of convenience — the languages taken as units for the analysis are those which are in some way defined already for the linguist. They might have emerged as the standard language for a region (as in the European languages or Classical Arabic), or they might be attested in a fairly unified corpus (as in Biblical Hebrew), or may be found in similar places and circumstances, as for the Ancient North Arabian languages (Macdonald, 2000, 2008a). Indeed, it is often not the linguist who does the choosing, but history.

This also means that the units of analysis in many studies of language history are also defined by social and historical processes, and not just linguistic ones. Castillian Spanish, for example, is not simply a variety which has certain linguistic features, it is also a culmination of various social acts that resulted in that variety being the dominant variety at the time of the creation of a nation state. This is not necessarily a bad thing; as LePage points out in the quote above — our unit of analysis should take into account how use of language is mediated through social groupings, and how social groupings are constituted in turn by language.

The problem however is one of comparability. The social and historical processes that led to the focusing and standardization of English are not the same as those which lead to the creation of French. Classical Arabic, Levantine Arabic and the *qeltu*-dialects of Arabic are not equivalent units to one another, and should not necessarily be treated as equivalent units to French, English, or Castillian Spanish. Whatever unit of analysis we adopt should be defined in such a way that different units are roughly comparable to one another.

Even when history has done the choosing, and a standardized variety is available for analysis, the fact that we can attach the same name to that variety at different points in time does not necessarily imply that there has been a single stable identical unit over time. The situation in English is illustrative:

Most of the Old English corpus is written in the Wessex dialect. because it was th[e speech of the west Saxon]... kingdom[,] the leading political and cultural force at the end of the ninth century. However, is is one of the ironies of English linguistic history that modern Standard English is descended not from west Saxon but from Mercian, ... the [ancestor of the Southern Midland] dialect spoken...in.. [and] around

London when that city became powerful in the Middle Ages. (Crystal, 1997, p. 29; quoted in Joseph and Janda, 2003, p. 19)

Here, we can see that Old English and the ancestors of modern Standard English are not a single linguistic entity in a simple diachronic relationship with one another; instead, they are different parts of a complex whole, a language which has had variation in the past as it has variation in the present, and where the dialect that made up that language became more or less influential over time. Moreover, the frequent references to social forces in that quote also shows that it is vital to understand the social conditions which influenced the rise and fall of different varieties, and not just their linguistic structures.

Another problem with attempting to treat the history of ‘English’ or even of the ‘Wessex’ or ‘South Midland’ dialects is that these dialects are not always the same unit through time. The case of English is quite clear — ‘Old English’ and ‘early modern Standard English’ are different entities, and would have been treated as such at different times throughout history. They would have merged into, or split off from, other varieties throughout their history, and so at a certain point it would be meaningless to speak of the Midland dialects as an unchanging and purely geographically situated entity. This is especially true of languages where speakers frequently migrate, as is the case with nomadic dialects of Arabic. Whatever unit of analysis we use to reconstruct a language like Arabic must be able to handle the changing links a dialect or language has over time.

Indeed, when researchers *do* assume a periodization of the language which implies direct linear descent from an attested older stage and a more modern stage of a language, the results are generally poor.⁶

⁶ This is not to say that direct lineal descent is could not possibly be the case, but rather that it

This has often been a problem in Arabic historical linguistics. For example, Garbell (1958) assumes that the phonology of the dialects in the modern Levant was essentially identical to, or directly inherited from, Classical Arabic, and therefore her periodization of the phonological changes that occurred in these dialects is quite flawed. She posits a merger of the /ð^s/ and /d^s/ ⁷ phonemes (orthographic ظ and ض respectively) sometime in the 11th to 15th century. In contrast, accounts that do not make this assumption argue that this was a very early merger, well pre-dating the Islamic conquests in the seventh century. Cowan (1960, p. 51), who takes a comparative method approach using only modern dialect data, considers that merger to be a feature of proto-colloquial Arabic, quite similar to the view held by Al-Wer (2004). J. A. Brown (2007) shows evidence that this merger had occurred within South Arabian speech communities by the fourth century CE, and that the grammatical tradition that emerged in the eighth century CE in Iraq (and which standardized what we now think of as Classical Arabic) was essentially dealing with a diglossic situation where contemporary spoken dialects had undergone the merger, but the high variety had preserved some unmerged forms.

Within Semitics, the assumption that Classical Arabic is near to, if not identical with, Proto-Arabic has also led to problematic work in reconstructing further back, a problem which is only recently being treated more seriously in Semitics (see for example recent work by Hasselbach, 2007a; and Pat-El, 2009). For example, the notion that all Arabic varieties have generalized /a/ as the vowel in their prefix conjugation verbs as stated by Faber (1997, p. 9) is clearly based on Classical Ara-

cannot be assumed without proof that a model of linear descent clearly accounts for the data.

⁷ The exact realization of this phoneme at a given time is a matter of some debate, with many scholars agreeing that at some point in the past, modern /d^s/ was a lateral fricative, but it is clearly widespread as an emphatic voiced dental stop in almost all dialects today (for rare dialect preserving a lateral realization, see Al-Azraqi, 2010)

bic, but even early grammarians noted /i/ vowels (so-called *taltala*) in variants of Classical Arabic and the variations in the prefix vowel is a frequent feature in contemporary dialect maps (Behnstedt, 1985, map 79; Behnstedt and Woidich, 1985, map 208).

Finally, the analytic unit of “language”, defined by the linguist, may also ignore the realities of use and speaker’s conceptions of how languages are related. Wolff (1959), discussed further below in Section 1.2.2, gives numerous examples of how linguists’ conceptions of how similar (structurally) two languages are rarely tallies with how they interact with those languages. For an example from the Arab world, the vast structural differences between French and Arabic do little to hinder Tunisians’ pervasive use of code-switching between the two languages, and a historical linguist treating them as separate due to their separate past would miss the fusional possibilities exploited by these speakers.⁸

A purely linguistic approach to difference, then, can miss how speakers relate to that difference, and the essentialist notion of languages defined by their accumulation of structural innovation will miss how languages can actually be used.⁹ On the other hand, we obviously cannot ask speakers of historical languages or varieties how they conceive of the boundaries of their languages. However, as discussed below in 1.3.3, the linguistic results of contact between varieties can offer valuable clues about how speakers conceive of the difference between them.

Thus, one of the major challenges faced by this study, and by the study of

⁸ This distinction between speakers’ and historical linguists’ perception of a situation seems similar to the distinction maintained by some linguists between e.g. Hock’s (1991) ‘recomposition’ (i.e. restoring etymologically correct divisions between parts of a compound that had become opaque) and ‘folk etymology’ (the same process, but etymologically incorrect.) To a linguist, these are different processes, but to the speakers they are one and the same.

⁹ See for example of the discussion on Ancient North Arabian and Arabic in Section 3.3.1.

similar languages, is what unit of analysis to employ in the process of reconstruction. Arabic has no clearly focused and named regional varieties equivalent to the Romance languages national varieties. The standard language, Classical Arabic, was probably quite distinct from spoken registers for some time prior to the Islamic conquests, and we've explored above why it makes a poor analytical unit. Indeed, even a focused variety (i.e. a 'language') is problematic as a unit of analysis, particularly in its focus only on linguistic form and exclusion of social valuation and use.

Instead, based on the discussion in this section, the unit of analysis for this study should satisfy a number of desiderata. The unit should be able to capture both linguistic structure and how that unit is situated socially. The unit should be sufficiently well defined that it is directly comparable to other, similar units. It should have some way to handle changes over time, and relationships which go beyond those of direct linear descent. It must also be able to reflect or capture in some ordered way the diversity inherent in language.

1.1.2 Genealogy

This study also differs from the comparative method in terms of its goal: Where the comparative method typically seeks to reconstruct the oldest possible ancestor (a proto-language), the goal of this study is to reconstruct a particular historical moment in the development of Arabic. I will refer to the 'ultimate ancestor' approach as the GENEALOGICAL APPROACH in contrast to the approach I will be using here, which I will term the SITUATED-HISTORICAL APPROACH.

The genealogical approach seeks linguistic evidence for a "period of exclusively shared prehistory, during which [dialects] are in close contact with one another, but not with the rest of the [family] (Hock, 1991, pp. 578-9)." Languages

which share this common prehistory are said to be related, provided we can exclude the possibility that some later diffusion of a change is what unites them.¹⁰ Indeed, the exclusion of so-called ‘areal features’ is one which consumes a great deal of the work in historical linguistics, especially Semitics where many of the Semitic languages seem to have been in intimate contact over long periods of time (e.g. the insistence of Al-Jallad, 2012, on separating areal from genetic features; but see also the discussion Huehnergard, 2011, which tries to reconcile ‘areal’ and ‘genetic’ signals).

However, the distinction between so-called areal and genetic features is not really sustainable. One of the primary observations of speaker-oriented (rather than language-oriented) approaches to language change (e.g. Croft, 2000; Enfield, 2003) is that the traditional division between internal and external change is rather illusory, especially between closely related varieties. An innovation must arise first in an individual’s use of language before it is taken up by different subgroups, before finally spreading throughout the population of ‘speakers of a language’ however that is defined. However, during the diffusion of that innovation, the boundary between ‘internal’ and ‘external’ change is not clear — from the perspective of those who haven’t changed, the change is ‘external,’ coming from a source outside the language that they are currently speaking (especially if we define language as being at least one innovation removed from their immediate siblings).

In this view, the genetic signal is really just an areal diffusion of a change which only spread across the ancestors of the ‘related languages.’ When operationalized in the genealogical approach, the distinction is really between changes

¹⁰ Or that they borrowed a given feature from a common linguistic source (hence the origin of contact linguistics in historical linguistics), or that feature is not an independent but similar innovation (i.e. a result of the lects’ typological poise, which is likely to be similar due to their family history. See Enfield, 2003).

which diffused at different times. If an innovation diffused at a sufficiently distant time in the past (or where we cannot distinguish its chronology relative to other changes) then it is a ‘genealogical’ signal, but if it is relatively, demonstrably more recent, then it is classified as an ‘areal’ signal.

For example, in his discussion of the development of West Semitic, Huehnergard (2011) distinguishes between West Semitic languages, such as Modern South Arabian and Ethiopic languages, and Central Semitic languages based on several genealogical features (see also Huehnergard, 2005). One feature, as an example, is the replacement of the imperfective form *yaqattal* ‘he kills’ found in Proto-West-Semitic with the Central Semitic form *yaqtulu* ‘he kills’. This replacement characterizes all of the Central Semitic languages, and so must have been a change that diffused throughout the communities of their ancestors.

In contrast, he also discusses features which don’t exclusively link all of the Central Semitic languages. Both the change of the final feminine marker $^{*}(a)t > -a$ and the reduction of final triphthongs *-aya* and *-awa* $> -a$: occurs only in the Central Semitic languages Hebrew, Phoenecian, Aramaic and Arabic; the northernmost and southernmost of the Central Semitic languages, Sayhadic and Ugaritic do not undergo these changes.¹¹ Here we have two diffusions of innovative features moving through the same group of languages, and these changes dictate the linguistic behavior of their descendant languages. However, since they are chronologically later, and apply to a different group of languages than the earlier change, they are not classified as a genetic signal, but rather as an areal diffusion (Huehnergard, 2011, pp. 267-268).

¹¹ Hebrew, Phoenecian, Aramaic and Arabic are not considered to form a genetic grouping in his analysis — there are other, older changes which place them in different subgroups relative to one another. Arabic descends directly from central Semitic, while the others are all Northwest Semitic languages, a subgroup of Central Semitic.

This classification makes sense if the question that is being asked is about ultimate origins: neither the loss of *-t* in final feminine marker, nor the resolution of triphthongs, was characteristic of the speech community that gave rise to the Central Semitic languages. It was only characteristic of a later stage of the language. However, if one is asking what different speech communities that a language has been a part of over time, the distinction between an ‘areal’ or a ‘genetic’ change is simply a difference in chronology, not a qualitative difference in the type of change that occurred or how it occurred.

The distinction could be likened to the difference between a birth certificate, and a *curriculum vitae*. If we wish to establish someone’s citizenship, a birth certificate is essential — we must establish a person’s genealogy to determine whether they are entitled to a given citizenship. However, when hiring someone for a job, a birth certificate tells us very little about someone except their ultimate origins; a CV is instead necessary to help us understand how the events in a persons life have prepared them for the position.

Indeed, a language’s recent history can be much more important than its pre-history. For those languages which fall off of our conception of a genetic family trees, such as pidgins and creoles, their more recent history is far more important to understanding their current states than is their distant pre-history. This is not just true for “exceptional” cases of contact such as pidgins. Shared features between languages which have different ultimate ancestors is extremely significant: It can tell us a great deal of information about how speakers conceived of the distinctions between those languages, it can tell us why kinds of relationships existed between groups speaking these ‘different’ languages and where they were located. In some ways, it gives us far more information on the recent history of these languages than knowing what linguistic features characterized their distant

ancestors.

1.1.2.1 Homogeneity of the past

Another important issue with the genealogical approach to reconstruction is that it generally aims to produce a homogeneous ancestor, a ‘proto-language’ which contains all of the reconstructed material, but no variation. Though this requirement has been discussed and criticized, and defended for a long time (see Dyen, 1969; Pulgram, 1959, 1961), it still forms an essentially assumption in much reconstruction. The assumption is essentially that a reconstructed proto-language is “the form of the parent speech (which is defined as a uniform language) (Bloomfield, 1933)”. This assumption is seen as necessary for the comparative method to operate correctly:

When we wish to employ the comparative method we are forced to make a potentially false working assumption: that the distinct languages which we are comparing trace back not merely to a single parent language, but to a single language FREE FROM DIALECT VARIATION (emphasis in original). (Hockett, 1958; quoted in Dyen, 1969, p. 502)

Dyen (1969) argues, however, that the necessity of reconstruction uniformity has not been clearly proven. While one can attempt to reconstruct what he calls a “maximally reconstructed proto-ideolect”, one can also attempt to reconstruct a more intermediate variety which does reflect dialect variation. Hock (1991, pp. 568-573) holds a similar view, distinguishing between proto-Indo-European and what he calls informally ‘late’ proto-Indo-European, whose dialect variation must have given rise to cross-cutting isoglosses found in branches of the Indo-

European language family (and expressed in a modified schematic unlike the traditional tree diagram).

The pressure to reconstruct a homogenized proto-language can force a researcher to make decisions which are in many ways arbitrary. There are many cases where there is no clear synchronic relationship between two forms, and therefore one cannot be said to derive from the other. To resolve this type of situation, the researcher may fall back on a principle such as ‘majority wins’, i.e. that “all else being equal [...] we tend to pick for our reconstructed proto-sound [or any proto-form - AM] that particular sound [or form - AM] in the correspondence set which shows up in the greatest number of daughter languages. (Campbell, 2004).” While all things being equal, we would expect the form that is most widespread at a given time to remain the more widespread form over time (hence this is an instance of Occam’s razor), linguistic history is rarely equal in this way, and it is rather the form spoken by more influential or powerful groups that tends to spread.

Moreover, as discussed above, there does not actually seem to be a true methodological need to exclude all variation from the proto-language; instead, it is closer to a convention of the field. In fact, the existence of two forms that are apparently equivalent in function and which cannot be related chronologically in any principled way should actually be a very strong indication of variation inherent in the ancestral community which the languages showing those variant forms descended from.

The question of homogeneity also relates to the issue of layers of diffusions and chronologizing changes raised in the previous section. Even (structurally) homogenized proto-languages don’t actually reflect a single chronological period, as H. Andersen (1996, p. 184) illustrates with Proto-Slavic:

Proto-Slavic [...] cannot be correlated with a single point in time [...] [the] substantival declension has a more archaic appearance than Proto-Slavic [verb] conjugation [...]. And when we consider the lexicon [...] it is very obvious that Proto-Slavic includes lexemes of widely different age.¹²

That is to say that even when proto-languages are structurally homogeneous, i.e. having no unstructured free-variation in their forms, they are actually chronologically diverse. Proto-languages are thus like the imaginary museum of the future described by the historian Felipe Fernández-Armesto (quoted in Joseph and Janda, 2003, p. 56) in which “diet coke cans [...] share with coats of chain mail a single vitrine marked ‘Planet Earth. 1000-2000 Christian Era.’”

Thus, the methods and units of analysis that are typically used in historical linguistic reconstruction are not entirely appropriate for understanding the history of a language like Arabic. Arabic is not a highly focused language, and it is not clear if a purely linguistic definition of Arabic would be as helpful as something which is more socially situated. Reconstructing an ancient, homogeneous proto-language does not tell us very much about the (relatively) recent history of Arabic. Arabic was probably differentiated from its neighboring languages in the sense used by historical linguistics for thousands of years prior to the Islamic conquests in the seventh century (Kitchen et al., 2009, see); however, the two or three hundred year period following those conquests is what defines to a great extent the current distribution of linguistic forms over space, though those forms developed prior to

¹² See Toulmin (2009, pp. 28-30) for similar statements regarding Indo-Aryan.

the conquests. What is needed is a framework which allows us to determine when changes arose in Arabic, how far they diffused, and then how we can relate this to social history in order to determine which changes had occurred prior to the Arab conquests.

1.2 From languages to speech communities

The preceding discussion shows that there are a number of desired traits that we need in a framework for reconstructing a particular time period in the history of a language like Arabic. Due to the extremely diverse dialects and no appropriate focused varieties, we need a unit of analysis that is more flexible than that of ‘language.’ This unit must be defined in such a way that different units can be compared with one another, and it should be able to capture the social context of language use, including speaker conceptions of language boundaries.

When tracing the history of Arabic in this framework, we also need to reconstruct something that is more historically situated than a proto-language. The reconstruction must also show the diversity of the language that existed at different periods in its history, and be able to capture the fact that linguistic varieties merge with and split off from other groups throughout their history. The framework of reconstruction should make use of what we know about the spread of features to understand the relationships between the languages which did or did not adopt those features.

The most promising framework which has those traits is that of reconstructing ‘speech communities,’ a framework pioneered by Ross (1997) who applies it primarily to Oceanic and Austronesian historical linguistics, but occasionally used by other scholars Toulmin (2009).

Ross (1997, p. 214) defines speech communities, following Grace (1996, p. 172), as “those people who communicate with one another or are connected to one another by chains of speakers who communicate with one another.” Thus, a speech community in this view is basically a social network in the sense of J. Milroy and L. Milroy (1985); L. Milroy (2008), which we will explore further in section 1.2.1.1 below. Similarly, Toulmin (2009, p. 23) focuses on “propagation networks” “the population of interacting speakers who participated in [a] change”, and Enfield (2003, pp. 11-14) focuses on social networks as determinants of speaker interactions. Linguistic changes are seen in these frameworks as bounded or influenced by the structure of the networks of speakers, so that a split in the social networks of people in a given speech community (i.e. severing of ties between members of subsections of the community so that they no longer communicate) is a sufficient condition to cause a linguistic split.

This mechanistic, social network centered approach does fulfill many of the requirements for our framework. It helps to situate language within a type of social space (that of the community, defined as people who communicate with one another), and it can capture a great deal of diversity — Graces’ definition of speech community deliberately makes no reference to how many languages are used in communicating, or how much variation exists in that communication. Speech communities are malleable and change over time, splitting, merging, differentiating in place, and this should have clear non-linguistic correlates also. Speech communities, since they are defined in the same way, should be roughly comparable to one another.

On the other hand, the social-network definition does little to capture how speakers conceive of the boundaries between linguistic varieties, which as will argue below in Section 1.2.1.2 is very much a determining factor in how individuals

and communities make use of language, which over time leads to language change.

Therefore, this section will argue for adding another component to our definition of speech communities: the allegiances of speakers to different speech communities. By adding this component, I argue, we are better able to capture the realities of language use. We are also better equipped to link what we know about speech communities from a linguistic perspective with extra-linguistic information, such as historical records, meta-linguistic traditions, and literary works. Since we are able to employ more sources of data, we can develop richer and more rigorous narratives about the history of the speech communities which we analyze than if we focused only on network structure.

1.2.1 Defining speech communities

1.2.1.1 Social Networks

Languages and more generally the notion of a ‘community’ are both typically defined with a reference to space. A language or dialect is characteristic of a certain region; a community similarly is a place where people interact face-to-face contact is necessary. For example, in the following anthropological (archaeologically oriented) definition of ‘community’, space is of great importance:

We see the community, in its simplest description, as the conjunction of “people, place and premise” to borrow a phrase from John Watanabe (1992). More specifically, it is an ever-emergent social institution that generates and is generated by supra-household interactions that are structured and synchronized by a set of places within a particular span of time. Daily interactions rely on, and in turn, develop shared premises or understandings, which can be mobilized in the development of common community identities. We do not neglect the spatial

aspect of the community because there must exist physical venues for the repeated, meaningful interaction needed to create and maintain a community, but we reject notions of community as solely a socio-spatial unit. Although this view does not imply co-residence of a community's members, it does require their frequent co-residence, at least for periods prior to the invention of technologies like the Internet and telephone that allow for frequent but physically distanced interaction.

There is, as the quote suggests, significantly more to a community than just space, and this is one motivation for including the notion of linguistic allegiance that we develop below (which corresponds to the notion of “premise” here). However, the notion of space is taken as characteristic both of communities and languages — both must be situated within physical space.

The difficulty with an appeal to a simple concept of space is that physical geography is not in the end a very good determiner of how people interact. A large state university is a good example of this — though at a given time, thousands of people are within a relative small space, possibly even within the same lecture hall, they are not likely to interact only with those nearest to them. Instead, they interact with people to whom they are linked by friendship or people from whom they need to demand certain services (including e.g. interactions with faculty but also staff, cafeteria employees, etc). The web of ties that each person has in the university is actually what determines who they interact with, rather than gross physical proximity.

The same is true of linguistic interactions more generally. Physical distance from the center of a linguistic innovation, for example, is a poor predictor of who will adopt that innovation — in many cases, an innovation will skip vast distances to the next place where it is adopted, the ‘hierarchical diffusion’ model that seems

to be characteristic of most innovations (Britain, 2008, gives a number of examples of ‘hierarchical diffusion’ on pp. 623–626; see also Taeldeman, 1995; Thomas Wikle and Guy Bailey, 1996; Thomas Wikle, Guy Bailey, et al., 1993). In Syria, for example, the use of the demonstrative *hayy* instead of *ha:di* appears to be diffusing from the largest urban centers (Damascus and Aleppo) to smaller urban centers (Palmyra, Deir Ez-zor, Qamashli City), but without affecting smaller or less important areas in between, so that Ar-Raqqa, between Aleppo and Deir Ez-zor is not affected, nor are most of the small villages between the two cities (Behnstedt, 1997, map 276).

What is happening here is the diffusion of change not over space, but over the social networks that connect people due to consanguinity or public interactions i.e. due to work or travel. Someone living in Damascus is more likely to travel directly to another city, such as Der Ezzor, for trade or work, or vise-versa, rather than a nearby village, and so the social networks involved are not diffused evenly over space. Hence, space per se is not as useful as social networks for understanding community, and specifically linguistic community.

Speech communities then must at their most basic be collections of social networks. The primary linguistic research on social networks was conducted by James and Leslie Milroy (for a recent overview, see L. Milroy, 2008), originally an idea taken from social anthropology. Replacing predefined variables such as class, social network research measures network strength, i.e. the number and strength of connections between speakers. The primary result of this research has been to differentiate between two types of networks: Close-knit networks with a large number (multiplex) of strong ties between members of the networks (density), and loose-knit networks with a small number of weaker, largely uniplex ties.

In the schematic in Figure 1.1 each node represents a person, and the lines between them the social links. Solid lines reflect strong, frequent interactions (as

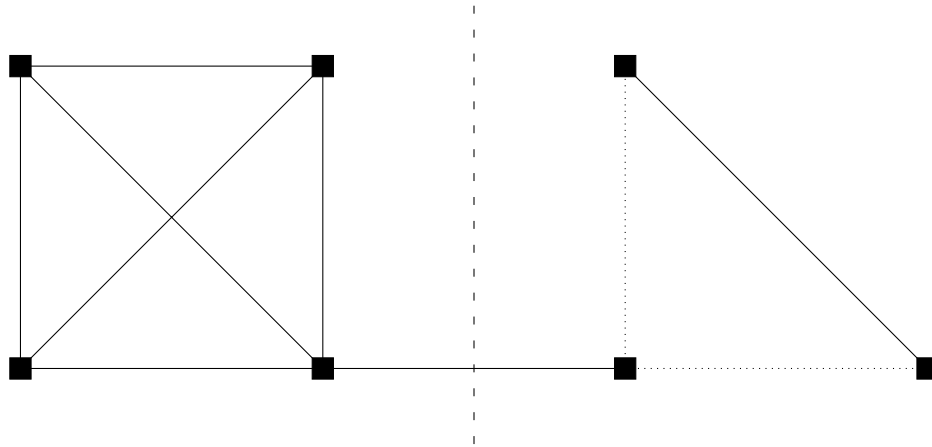


Figure 1.1: Two connected networks, a dense one on the left, and a loose-knit network on the right.

in close friendship), while dotted lines reflect weak, more occasional interactions (acquaintanceship, purely functional interactions). There are two main networks here, separated by the dashed line. The leftmost network is more multiplex, with everyone member of that network knowing everyone else, and dense, with strong connections between all members of the community. The network on the right is less dense, with not all the people in the network interacting with one another directly (i.e. less multiplex). Moreover, the connections in the network on the right are generally less strong (i.e. less density). However, since the two networks are connected, though by a single pair of speakers, they also form a larger social network, which means that social networks are recursive and can potentially be either large or small.

Network structure plays a large role in how changes diffuse over the network. Close-knit networks tend to act as insulators from change and tend to reinforce group behavior — in the words of L. Milroy (2008, p. 563), “close-knit network

structure supports localized linguistic norms and resists change originating outside the network.” Speakers with weaker ties, however, are not as strongly bound by the norms of the close-knit groups, and are therefore more likely to bring innovations into a community from the outside. The weak ties between Protestant and Catholic women living in different parts of Belfast separated by a physical barrier of barbed wire and concrete allow changes to diffuse between the two communities. Similarly, hierarchical diffusion of features may be driven by the many weak ties between speakers in two areas. Trudgill (1986, p. 54) gives the example of tourists and soccer fans traveling from London into the Norwich, bringing with them linguistic forms which help to alter the linguistic behavior of the Norwich community. These London English speakers don’t become deeply embedded in the local social networks, but do interact on a weak commercial and perhaps personal basis with the locals, i.e. creating many weak ties.¹³

On a broader level of analysis, J. Milroy and L. Milroy (*ibid.*, pp. 377–380) argue that the relative conservatism of Icelandic versus English (as well as the greater dialect diversity of English) is a result of the greater mobility and weaker social ties of English speakers as compared to speakers of Icelandic who had stronger group cohesion. A similar argument has been proposed by Penny (2004, pp. 63–67) to explain the spread of certain linguistic features in Castilian Spanish¹⁴— for an

¹³ Of course, since a group outsider is unlikely to lead linguistic change, J. Milroy and L. Milroy (1985) posit a difference between INNOVATORS who are marginal to a social group, and EARLY ADOPTERS who are central members of a group and who typically provide a model for other members of that group. Innovators, in their view, are responsible for moving changes around by virtue of their weak ties, while early adopters’ weak ties to the outside world give them access to the innovations, while their strong ties and position as a leader of style within their groups generally, allow them to diffuse those changes with relative little resistance. The level of analysis that we will be able to discern in a historical linguistic study will not necessarily be able to reach the level of individuals, and so we will not go into this topic in depth.

overview of similar historical examples see Conde-Silvestre (2012).

Network structure is similarly important in understanding the history of Arabic: Groups that are truly transhumanist nomads will likely have very dense in-network ties, and very few out-network connections, and thus changes arising outside of their networks are unlikely to be adopted by these groups. Groups that have a more sedentary lifestyle are likely to have somewhat looser internal ties, and a greater number of weak, group external ties as they interact with other nearby groups. The resistance of nomadic groups to innovations diffusing from urban centers, as well as their long-distance mobility causes the apparent schism between Bedouin and sedentary dialects, as discussed in Section 2.5.

1.2.1.2 Identity and Allegiance

Social network position does not, however, determine speaker behavior. Fought (2012, p. 286) discusses the example of African American speakers from Henderson (1996) who, though integrated into the social networks of the Philadelphia white community, do not share in their use of language (specifically raising of short /a/), as they felt that they were treated by the white community as not only different, but inferior. Their refusal to join in with the behavior has little to do with their social network position or links; instead, based on how they and others perceive their identity, they make a (semi-conscious?) choice of who to display allegiance to linguistically. Thus, we cannot conceive of a speech community as a network composed of dumb nodes, but rather as a collection of human beings with individual agency, and in this section we will argue that how speakers negotiate

¹⁴ However, changes in social network ties were not the only factor. As Penny (2004, p. 66) notes, these changes also included social changes in the openness to other cultural modes of practice, including Moorish culture, as well as innovations against traditional jurisprudence, social structures and even script.

their identity and allegiances linguistically plays a major role in language change.

Indeed, attempts to define speech communities based solely on their network structure often show a ‘leakage’ from other facets of language use, particularly speakers’ understandings of languages and the borders between them. Ross (2005) tries to formulate a diagnostic approach to language based entirely on network structure, but throughout the paper, he is forced to make reference to speaker attitudes. Though he draws his framework from H. Andersen (1988), who makes a distinction between endocentric (bound by tight bonds of linguistic solidarity, Ross’ ‘tightknit’), or exocentric (having less communal linguistic solidarity, Ross’ ‘looseknit’), Ross (2005, pp. 179-180) prefers to see these as consequences of network structure, which co-varies with speaker attitude. However, he nonetheless makes reference to speaker attitudes throughout the article, even when they are not simply the consequences or causes of network structure: He refers to the notion of “emblematic” linguistic forms (which must be seen as such by speakers, regardless of their network position) or of languages treated as “variants of each other rather than as separate languages (p. 191)” — a change not in network structure, but in speaker conceptions of what boundaries exist between languages. As we discussed above, social attitudinal factors can be entirely independent of network position and structure, and it is only by acknowledging the different contributions from both attitudes and network structure that we (or Ross) are able to flesh out a clear picture of how speech communities change over time.

This argument is elegantly summarized by the dialectologist Walt Wolfram:

[D]ialect adoption is not a simple matter of who you interact with under what circumstances - it’s a matter of how you perceive and project yourself - much more capturable in cultural identity schemes than interactional reductionism. (cited in Hazen, 2000, p. 126)

Similarly, the anthropologist A. P. Cohen (1985) states that “in seeking to understand the phenomenon of community we have to regard its constituent social relations as repositories of meaning for its members, not as a set of mechanical linkages.”

The crux of this linguistic behavior lies in the human desire for group belonging. Social psychology, particularly the work of Henri Tajfel, has found that people form groups extremely quickly, and in experimental settings it matters little what criteria is used for group assignment — even assignment by random number was sufficient to create group behavior that favors in-group members and disfavors out-group members (for discussion and citations see Edwards, 2009, pp. 25-27).¹⁵

Groups and communities are typically defined relatively by constructed, socially conditioned boundaries and not some absolutely defined criteria. These boundaries, as argued by A. P. Cohen (1985) based on the review of anthropological case studies, are relative, in that they are only defined in relation to other communities. They are delineated in various ways, but often by the deployment of symbols which are seen as representative of the community, such that “people’s experience and understanding of their community [...] resides in their orientation to its symbolism.” (*ibid.*, p. 16).¹⁶ This symbolism can be almost anything:

Since the boundaries [between communities] are inherently oppositional, almost any matter of perceived difference between the community and the outside world can be rendered symbolically as a resource

¹⁵ Anyone who follows international politics will be aware of the fact that this is also largely true outside of the psychology laboratory.

¹⁶ Note the similarity here to Labov’s definition of a speech community as being unified not by its linguistic behavior per se, but rather by “participation in a set of shared norms,” that is to say a shared understanding of the symbolism of linguistic material (Labov, 1972, p. 120).

of its boundary. The community can make virtually anything grist to the symbolic mill of cultural distance, whether it be the effects upon it of some centrally formulated government policy or a matter of dialect, dress, drinking or dying. The symbolic nature of the opposition means that people can ‘think themselves into difference.’ The boundaries consist essentially in the contrivance of distinctive meanings within the community’s social discourse. They provide people with a referent for their personal identities. Having done so, they are then themselves expressed and reinforced through the presentation of those identities in social life. (A. P. Cohen, 1985, p. 118)

Within the linguistic literature, this process of deploying linguistic material, i.e. speakers using language to ‘speak themselves into difference’ is referred to by LePage and Tabouret-Keller (1985) as “acts of identity.” These acts of identity allows the individual to:

create for himself the patterns of his linguistic behavior so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished (ibid., p. 181)¹⁷

These acts of identity go beyond individual preference. They are one of the important forces that maintain the integrity of a community::

¹⁷ The changes that speakers make linguistically to distinguish themselves from their neighbors can be quite extreme. Thomason (2007) gives the an extreme example of a community of speakers in Papua New Guinea who, concerned about the similarity of their dialect to those nearby, switched their masculine and feminine anaphoric agreement markers to be the opposite of their neighbors.

[T]he symbolic repertoire of a community aggregates the individualities and other differences found within the community and provides the means for their expression, interpretation and containment [...] It continuously transforms the reality of difference into the appearance of similarity with such efficacy that people can still invest the ‘community’ with ideological integrity. It unites them in their opposition, both to each other, and to those ‘outside.’ It thereby constitutes, and gives reality to, the community’s boundaries. (A. P. Cohen, 1985, page number?)

This kind of symbolism doesn’t act simply as positive definition of in-group boundaries, it can also influence how one conceives of outgroups by defining in-groups negatively by their lack of linguistically symbolic features. For example, in research by Niedzielski (1999), Michiganders were better able identify “Canadian raising” of vowels (present also in Michigan dialects) when told they were listening to a recording of a Canadian (really a Michigander). When told they were listening to the speech of a Michigander, however, they were unable to identify the same feature with precision. Here, the in-group is defined by a lack of stigmatizing, symbolic features, while the out-group has those features (Preston, 2013, reviews the literature on how allegiance, ‘regard’ in his framework, influences perception).

On the level of different languages, Wolff (1959) documents numerous examples of language pairings where the speakers of those languages have significantly different intuitions about intelligibility than (non-native) linguists working on those languages. These intuitions are not always symmetrical either: One group considers another’s language unintelligible, while the other group thinks both of the languages are similar. Considering a language intelligible or unintelligible therefore seems to be more of a linguistic attitude than a fact of language

structure, though of course linguistic structure will influence the type and degree of influence of languages on each other. Moreover, as long as speakers view another language as opaque, they are unlikely to draw on its linguistic material in any meaningful way.

To illustrate the role of identity and allegiance in language use, we can look at the new-city dialect of Amman. Here, third generation speakers are able to employ different variants of the Q variable to mark inclusion in various social groups, or to become more or less similar to an interlocutor (Al-Wer, 2007). Men are generally expected to use the /g/ realization when interacting with each other, and but are expected to use the female /ʔ/ variant when flirting with women. However, group identity can trump this, so that men of Jordanian, Bedouin descent use /g/ even when speaking to women, and in my own experience some Jordanian, Bedouin women will use /g/ as well, their ethnic identity trumping their gender identity.¹⁸ The cumulative acts of these speakers, even if there are some exceptions, create strong community norms, and even the deviations from those norms are evaluated based on the symbolism of the dominant uses of those symbols.

It is important to note that the group identities that are indexed by these linguistic markers are not *linguistic* identities per se. They are general identities, based on ethnic, or gender, or political bases, but which are marked using linguistic material. That is to say that speakers' identities are normally extra-linguistic, but they are marked linguistically. They are also usually marked in many other ways, as explained in the quote from Cohen above, but obviously for the study of the history of language, we are primarily interested in the linguistic symbols that speakers use.

¹⁸ I have found the same to be true for some Palestinian men also, whose identity as Palestinians is more important than sociolinguistically demonstrating their masculinity.

That is not to say that every piece of linguistic material used by a speaker is *consciously* symbolic of their identity. Labovian sociolinguistics (cf. Labov, 2001, pp. 196-197) recognizes three levels of awareness of variables which begin as internal changes ('change from below'): INDICATORS, which are used by particular social groups, but which are below the level of social awareness. Once they acquire social recognition they become MARKERS and are detectable by showing up in style shifting behavior in Labovian sociolinguistic interviews. Finally, if they are sufficiently socially relevant they become stereotypes, the topic of meta-linguistic comment and judgment. In Labov's characterization of markers and stereotypes, they are typically evaluated negatively, especially stereotypes, though a more neutral term would be useful for those linguistic features which are simply seen as marking the boundaries between communities, for which I will use the traditional term SHIBBOLETH.¹⁹

All three of these designations, however, presuppose some social association, that is to say even indicators, though below the level of perceptual salience, are associated with some group (in Labov's research, typically a social class). Any feature that arises or becomes popularized will spread to the approximate limits of a speech community. At first, it will probably spread only to a small scale group — a circle of friends, perhaps — and then it can spread more widely, where it might be indicative of, for example, a social class or neighborhood. Some changes will spread even more widely — through the entire speech community that conceives of itself as speaking a single language, for example. In that case, we would have a classical example of language change.

The vital point here, however, is that at every stage of that diffusion, the

¹⁹ Indeed, the general concept of shibboleths bears more research, as it has tended to be ignored in modern linguistic frameworks. See e.g. Kristiansen (2003, pp. 78-80).

feature is indicative of a speech community. At every stage, the diffusing feature is ‘grist to the symbolic mill’ of the speech communities involved, whether they use it to mark in-group or out-group status.²⁰ Therefore, at every stage in the diffusion of a feature, conscious or unconscious notions of identity and group allegiance dictate an individual speaker’s, and the communities in which they are embedded’s, use of language.

1.2.1.3 Speech Communities: Networks, Allegiance, Size

What I have shown here is that the social network approach and the acts-of-identity approach complement each other and help us better understand how changes in social structure and in social allegiances can impact language. Social networks situate a speaker within a social-communicative space, while social allegiances situate a speaker within an ideological space. Both a speaker’s network position and their allegiances will influence their language use, and changes to either will change, over time, how a speaker uses language.

We can therefore define a speech community as follows:

A SPEECH COMMUNITY is a group of speakers who are connected by a social network, and who share common allegiances.

The definition makes no reference to a common language. A speech community can be monolingual or multilingual, diglossic or polyglossic. What is important is that the speakers of those potentially different languages are unified by shared allegiances and part of a connected social network.

²⁰ Of course, as a feature becomes characteristic of increasingly larger speech communities, its role in distinguishing *internal* divisions is diminished — if everyone in a given speech community uses a given feature exclusively, then it cannot by definition mark difference.

The role that speech communities plays in the historical transformation of a language is that at every stage in the diffusion of a feature, the limits of diffusion mark the boundaries of some speech community. The mechanism behind this, as discussed above, is the process of language use in which speakers deploy linguistic resources to mark their identities, while the physical structure of a social network determines the extent of diffusion.

This definition of speech community is necessarily recursive: it could describe anything from a high school clique to a transnational linguistic area. Speech communities, like all communities, contain within themselves overlapping communities which themselves may contain other communities (A. P. Cohen, 1985; Croft, 2000, pp. 166-173; P. Brown and Levinson, 1979). The degree of heterogeneity will increase as the scale of the community increases, but it will still be bound together by (possibly increasingly vague) senses of belonging that allow for (and enforce) convergent linguistic behavior. This recursion is at the heart of historical linguistics, and captures the observation that changes typically diffuse through smaller-scale communities (e.g. a social class) then larger ones (a regional dialect) and finally throughout the entire community using a given language.²¹

However, in order to be able to compare speech communities to one another, it is helpful to define three sizes of speech community. The first is the micro-level which is typically the subject of social-network and communities of practice

²¹ As discussed previously, this is generally not so simple. In historical linguistics, it often appears as though a single innovation traveled throughout a single linguistic community, but it is very rare that we find synchronic examples where this is true — there is almost always some social group or marginal area where the change has not entirely diffused (consider the distribution of the word *ʃa:f* ‘to see’ in Arabic, where only a small number of marginal dialects do *not* use that word, though there is much evidence that it was not always so widespread.). The contradiction is the result of the pruning nature of history, where those marginal groups or dialects eventually disappear from the historical record — history creates an illusion of linguistic unity.

research, reaching the level of tens of individuals at most. This level corresponds to Croft's (2000) *DEME* which taken from biology refers a group of organisms in sufficient proximity to one another that mating is equally likely between any two appropriate members.²² In linguistic terms, this corresponds to "a group of people who are most likely to communicate with each other, and not so much with those outside of the network" (Croft, 2000, p. 20). This level of analysis is important to recognize in terms of its contribution to the larger processes that shape language, as detailed above, though it is not of great use for understanding the broader history of language. I will refer to communities of this size as SMALL-SCALE COMMUNITIES

A more appropriate level of analysis for historical research is the one that corresponds more closely to the size of a large dialect or a language, a speech community numbering in the hundreds and thousands in the pre-modern era and possible as much as millions today. Though of course there will be more internal heterogeneity than in small-scale communities, the important factor is that they be a part of the same social networks and view themselves as a speech community and that this inward facing orientation is the mechanism that maintains a certain linguistic homogeneity with respect to linguistic behavior. I will refer to this sized of community as a REGIONAL SPEECH COMMUNITY. This level of analysis is the focus of historical research and it seems unlikely that tools such as comparative reconstruction would uncover significantly smaller groupings.²³

²² Though like a community of practice, hierarchies may skew those probabilities

²³ Croft (*ibid.*, p. 19), again drawing on biological terminology, calls this level of analysis a GEOGRAPHICAL RACE. This originally referred to a subpopulation of a species defined geographically and with mild divergence in their physical structure but without full-blown speciation. In his terms, a geographical race corresponds to a geographical dialect "defined geographically, slightly divergent structurally, but not enough to prevent communication [...] or to provide a different sociolinguistic identity." Croft's definition, however, is tied to geography, while my definition is

Speech communities can be even larger than this, as exemplified in the notion of *SPRACHBUND* or *LINGUISTIC AREA*. Here vast areas are united into a large-scale, multilingual speech community where frequent use of languages leads to convergent linguistic features that exclude those outside of the area. Examples of this have been discussed widely in the literature (Dahl, 2001; Donohue, Wichmann, and Albu, 2008; Dryer, 1989) and include the Balkan sprachbund (Adamou, 2012), Mesoamerica (Campbell, Kaufman, and Smith-Stark, 1986) and the Indian subcontinent. Though much of the research has been focused specifically on how these areas can be differentiated linguistically, with surprisingly little attention paid to issues of identity and community (see for example Campbell, 2006), it seems that some sort of cultural sense of similarity, and not just geographic proximity, is necessary for a language area to develop. That is to say, language areas form as a result of being part of a larger speech community, which I will refer to as *AREAL COMMUNITIES*.

Indeed, non-linguistic cultural objects tend to precede linguistic accommodation throughout an areal community. Beier, Michael, and Sherzer (2002) argue that in lowland South America, a ‘discourse area’ exists wherein both quotidian and ritual discourse registers are shared among numerous cultures comprising a very large area. They note that a sub-area of this larger discourse area, the Upper Xingu, appears to be forming a linguistic area as well, though many of the languages and cultural groups involved have only been active in the area for about one hundred years. Discourse forms seem to be spreading faster than linguistic forms, from which they conclude:

This makes plausible the hypothesis that the diffusion of discourse

tied instead to social networks, since social networks may be widespread geographically.

forms are the primary means of diffusion of language-related phenomena, including strictly linguistic forms. In short, linguistic forms are diffused by means of the discourse forms that contain them, which can subsequently become adopted into everyday speech. (p. 137)

Campbell (2006, p. 558) similarly notes that the Mesoamerican linguistic area has a series of shared traits that might be considered part of what he calls “the ethnography of communication”, such as ritual language that involves paired semantic couplets, whistle speech and certain uses of pronouns for polite address. It seems nearly impossible that these groups would acquire such potent cultural and linguistic symbols from their neighbors without considering themselves to be in some way linked as part of a larger community, especially when these symbols are typically the very types of things which mark boundaries per A. P. Cohen (1985).

An analog within (and beyond) the Middle East is the cultural area composed of those smaller speech communities which share Islam as a religion. These communities speak a huge number of different languages, but their theological language is largely similar, drawing primarily on Arabic. Arabic terminology may spread even further, to function words (as in Turkish borrowing of *ve < wa* ‘and’) or to theophoric phrases such as *ʔin ʃa:ʔ alʔlʔa:h* ‘if God wills’ which are used in pragmatically similar ways throughout that cultural area.

1.2.2 Uniting the Linguistic and Extra-Linguistic

One of our primary desiderata for a new framework is that it be able to better integrate linguistic and extra-linguistic information, and this is indeed one of the primary advantages of a speech community approach. More specifically,

since we define speech communities in terms of two components, network structure and group allegiance, we can relate those components to changes within the community itself and therefore changes in language.

1.2.2.1 Changes in Structure and Allegiance

For example, if a speech community breaks apart (via migration, etc), we do not expect features to spread between what are now two separate speech communities, as they no longer sufficiently linked to one another by network ties for linguistic changes to propagate between them. Similarly, speech communities which are physically geographically separated from other communities around them tend not to participate in the spread of features. Mountain dialects tend not to participate in the diffusion of features of lower-land dialects, so they tend to have significantly different linguistic behavior from dialects that are nearby as the crow flies; the Arabic dialect of Jabal Rāziḥ, for example, may be quite divergent from nearby dialects at least partially due to the difficult terrain of the area (Watson et al., 2006a).

Similarly, in Section 2.4, I argue that the often groups referred to as a single ‘tribe’ in the pre-Islamic era did not form a cohesive linguistic community, as the social network connections between its members were likely too weak, in spite of a sense of allegiance. This changed in the post-Islamic *amṣār*, newly constructed cities, where tribal identity was exaggerated as the tribal unit became a unit of administration and housing was determined by tribal affiliation.

Physical distance and the attending linguistic fragmentation, however, can be overcome by a sense of allegiance to a larger speech community. This is not that unusual; Croft (2000, p. 16) refers to languages in this type of situation as POLYTYPIC LANGUAGES, “linguistic varieties that are structurally so diverse that lin-

guists would characterize them as different languages, yet their speakers perceive them as dialects of the same language.” The fact that Arabic-speaking communities see themselves as united by Classical Arabic allows these dialects to make nonce and more general borrowings from the Classical register. The allegiance to the Arabic community makes Classical Arabic available as a source of linguistic material and unites dialects which have diverged significantly and which at various lower-levels perceive themselves as different communities. Communities which no longer perceive themselves as a part of this language community (e.g. Maltese) no longer make use of this resource. Indeed Maltese is often used in studies on Arabic precisely because Maltese speakers have not conceived of themselves as belonging to the Arabic-speaking community for some time.²⁴

Similarly, changes in political allegiance often result in shifts in allegiance (and sometimes network structure), and hence in language use. As Croft (2000, pp. 18-19) notes, a speech community approach to language change predicts that a social divisions will lead to linguistic divergence, an observation not captured by a language centered approach. Croft (*ibid.*, p. 16) refers to languages which have were previously part of a single speech community that then underwent a social split as SIBLING LANGUAGES, that is:

Two linguistic varieties that are structurally so similar that they are considered to be ‘dialects of the same language’ yet are perceived by the speakers - or at least by one group of speakers - as distinct languages.

²⁴ Brazilian and Peninsular Portuguese (or for that matter, Spanish) are another example of structurally divergent languages that are still conceived of as different aspects of the same speech community (Clements, 2009, p. 37).

An example of this is the recent political division between German and the Netherlands, which has led to sharp linguistic divergence between what had previously been largely similar and often mutually intelligible languages. The establishment of the two countries reshaped the allegiances of speakers within their borders. The result has been a divergence from the usage of nearby, but now ‘foreign’ villages towards a convergence on national (rather than local) norms, reflecting a change in speaker affiliation, as shown in Figure 1.2.²⁵

How a group is situated with respect to other groups may also govern how language use changes. Mufwene (2004) differentiates between two major types of colonial domination (whether in modern or pre-modern times.) Exploitation colonies are those where a relatively small number of colonialists live, and their primary focus is extracting resources from that colony. Residents of these colonies don’t necessarily develop a shared identity, and certainly no allegiance to the colonizers. Settlement colonies, on the other hand, are intended as permanent places of residence. Settlement colonies tend to develop stronger internal allegiances and identities, which can be separate from that of the origin of the colonists. The linguistic results of this is that in settlement colonies, linguistic homogenization tends to happen much more rapidly than in exploitation colonies. As discussed in Section 4.1.2, the shift from Coptic to Arabic in Egypt may have begun only after Egypt stopped functioning as a settlement colony for the Abbasids, and because a settlement colony of the Fatimids (originally hailing from Tunisia.)

These examples illustrate the value of modeling language change in terms of speech communities — while obviously languages are diversifying constantly,

²⁵ These borders are rather porous, so speakers can still move between the countries, but there may have been a shift in network structure also, for example as children attend national schools appropriate to the side of the border that they reside on and thus form stronger network connections with their countrymen rather than those across the border.

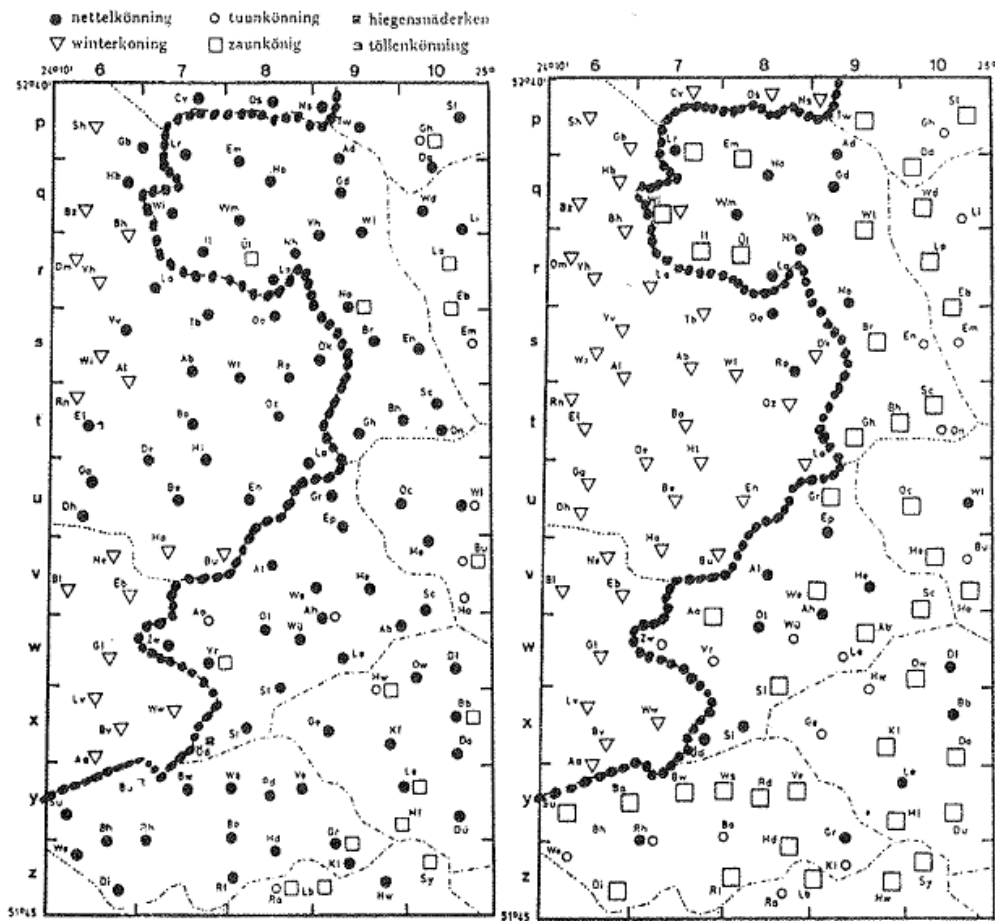


Figure 1.2: Realizations of ‘wren’ on the German-Netherlands border near Enschede/Rheine. 1940 left, 1975 right. From Auer (2002, figure 10) originally in Kremer (1990)

we can point to specific historical events that lead to the split of (speech) communities, and hence to the separation that leads towards the divergence of languages, rather than simply viewing it as an automatic process. In the terms of Croft (2000, p. 199), the ideological barriers between speech communities are “isolating mechanisms”, which like geographical barriers, lead to gradual differentiation.

In fact, an abrupt shift in speaker allegiance may lead to more rapid changes than the slow action of centuries of minor changes or migrations. The divergence of German and Dutch seems to be happening quite rapidly, within a time-frame of a century, while some linguistic areas, like that being formed in Upper Xingu (Beier, Michael, and Sherzer, 2002), seems to have been converging in a similar time-span. In contrast, Joseph and Janda (2003) cite Bynon, 1977 as suggesting that for the general application of the comparative method to diverging dialects some four to five hundred years of differentiation is needed for good results. That is to say, if there is a clear, deliberate shift in ideological orientation, it may cause faster differentiation than the small, gradual shifts that we think of as the heart of historical linguistic change.

1.2.2.2 Diagnosing Speech Communities

Another benefit of the speech communities approach advocated here is that, by making the notion of group allegiance and identity central to the notion of speech communities, we are able to exploit a wider range of meta-linguistic and literary texts. These types of texts can have extremely rich information on attitudes and even linguistic attitudes, and may help us develop a better understanding of how groups related to one another.²⁶

Literary and grammatical sources (the line between the two often being thin in the Arabo-Islamic tradition) can also help us understand earlier inter-group relations, and hence how their allegiances might have affected their linguistic be-

²⁶ This relates to what Joseph and Janda (2003) call ‘informational maximalism,’ that is the “utilization of all reasonable means to extend our knowledge of what might have been going on in the past,” specifically the prior states of languages. The speech community approach provides a principled way to integrate many kinds of extra-linguistic information that otherwise would be hard to fit into a linguistic, essentialist approach to language history.

havior. Bashear's (1997) *Arabs and Others in early Islam* draws on a variety of literary sources, including the Qur'ān, to understand what groups existed within the "Islamic" polity and how they are related to each other — for example, he finds that in spite of a later tradition which extols Bedouin tribal lifestyles, early depictions of Bedouin were extremely negative and cultural prestige likely rested with settled peoples, though both groups were "Arab" (a widespread pattern throughout the region and in different eras, cf. Shaw, 1982). This would suggest that settled peoples were unlikely to consider themselves part of the same speech community as nomads, and so we can expect significant linguistic divergence between those two populations in the pre- and early Islamic era at least.

Grammatical sources can also provide us with more than just information on what uses existed at a given time. For example, Sībawayhi, in his section on phonology, does not simply describe which sounds exist, but rather groups them into acceptable and unacceptable variants:

وتكون خمسة وثلاثين حرفا بحروف هن فروع وأصلها من التسعة
والعشرين وهي كثيرة يؤخذ بها وتستحسن في قراءة القرآن والأشعار
وهي: [...]

وتكون اثنين وأربعين حرفا بحروف غير مستحسنة ولا كثيرة في
لغة من ترتضى عريته ولا تستحسن في قراءة القرآن ولا في الشعر
وهي: [...]

There are 35 variant phonemes which are based on the 29 acceptable phonemes, they are frequent and usable as models, and acceptable in reading the Qur'ān and poetry, and they are [...]

There are 42 sounds which are not acceptable, nor frequent in the language of those whose Arabic is acceptable, and they are not acceptable for Qur'ān or poetry recitation. They are [...]

This passage does more than tell us what realizations for shibboleths were present in his time, it also gives us a clear view of the relative prestige of certain realizations. Prestigious forms often reflect prestigious groups, and if we relate a reconstruction of earlier speech communities with this portrayal of the relative prestige of different forms, we could even reconstruct the power relations that likely existed between those speech communities, as viewed of course through the lens of an urban grammarian in southern Iraq.

1.3 Towards an archaeology of speech communities

The previous section explored how we can utilize extra-linguistic information to determine which speech communities existed at a given time, and how they related to one another. In this section, we explore what primarily linguistic data can tell us about the composition of speech communities, and their relationships to one another. The nature of our data and the goals of our reconstruction mean that we will largely be dealing with ‘dialect’ sized, regional speech communities, though occasionally we will be able to diagnose communities of much larger sizes, akin to a linguistic area.

The focus in this section, for purposes of brevity, is going to be on contact between languages which do not result in the creation of pidgins, creoles or mixed languages. Accounting for these processes within the framework of speech communities is certainly reasonable,²⁷ but in order to keep this section short and most useful for the dissertation, I will not try to cover these types of language events since they are rare in the history of the Arabic language.

²⁷ For example, creolization of a pidgin can be described as a result of the fusion of speech communities which had previously remained separate.

1.3.1 Alignment Events and Innovation

The essential notion that underlies our analysis here is what I will refer to as ALIGNMENT EVENTS. An alignment event is when a speech community (of any size) adopts or focuses a form bringing it into alignment or out of alignment with other speech communities with which it is in contact.²⁸ Failure to participate in an alignment event may also be diagnostic of speech communities, as in dialect enclaves where all surrounding areas have realigned to a common linguistic practice, though other diagnostics would be necessary to ensure that it the alignment simply hasn't happened *yet*.²⁹

In most circumstances, the alignment event will take the form of an increasingly large speech community aligning itself to a linguistic innovation. This would correspond to the innovation and diffusion of a feature as treated in sociolinguistics and historical linguistics. For example, the /ʔ/ realization of the variable Q is a clear innovation which has already spread through the urban centers of the Levant and Egypt, but is continuing to diffuse through speech communities in the Levant (see for example the areas around Aleppo in, Behnstedt, 1997, map 9)

An important distinction here between the normal view of historical change in languages and the alignment event approach is that the material that is employed by a speech community engaged in an alignment events need not be based solely on innovations but may also take as “grist for their symbolic mill” (per A. P. Cohen, 1985) any linguistic material.³⁰ Stated more simply, *linguistic features* are

²⁸ These changes would be best defined in relation to the speaker's (in the changing speech communities) conception of the other speech community, which may or may not be a real community. This allows us to account for diglossic changes, where the high language can be seen as representing some sort of (non-existent, and possibly historically non-existent) speech community.

²⁹ In a longer view, anything could be seen as not having happened yet, so there obviously needs to be bounds to this concept so that we do not end up in a logical game of *reductio ad absurdum*.

³⁰ Indeed, the distinction between innovations and retentions is, at a certain level of analysis, rather

what diffuse through a population, not only features where are ‘innovative’ relative to some historic past.³¹

However, more important is the fact that speech communities may use features which are themselves quite ancient (relative, e.g. to their language family, or even the history of their language) to distinguish themselves from other. In Martha’s Vineyard, the centralization of the two variables AU and AI was ‘innovative’ and ‘retentive’ respectively, but they showed similar usage to distinguish those whose identity was bound to the island and those who had a more outward orientation (Labov, 1963). The centralized realization of AI had been traced back to the sixteenth century in England, and was at one time widespread and even potentially a prestige variant through the nineteenth century in the US. However, in the case of Martha’s Vineyard, the status of the centralized realization of this variable is only relevant in contrast to other, adjacent social groups which have lost the

arbitrary. In a very strict (and rather reductive) view, what is considered retentive should reset with almost every generation, and only developments that occur within that generation could truly be seen as innovative.

³¹ The notion of ‘alignment events’ is quite similar to Toulmin’s (2009) PROPAGATION EVENT, which he defines as when “an innovative feature is propagated through a web of interconnected networks of speaker interaction. (p. 23)” Propagation events are limited in their range (geographic, social, and temporal), and Toulmin (2009) sees the task of the historical linguist to determine what propagation events occurred at that times and in what sequence. His notion of propagation event is limited by the typical historical linguistic caveats that it must be innovative, vis-a-vis a “more historically distant stage (p. 35)” and the researcher must exclude the possibility that two identical (in linguistic form) propagation events occurred (rather than one) or that the propagation event originated in a similar source (due to the “ecology” of the language area, i.e. the shared diglossia between local languages and Hindi in the area he studies).

His approach differs from ours in our emphasis on the social nature of these changes, and his emphasis on innovations. While many alignment events are necessary innovative with respect to the immediately prior historical state of the speech community, Toulmin still tends to depend on innovations defined from a larger frame historical perspective. An alignment event is a detectable change in behavior on the part of speakers, which is necessarily innovative in some sense, but our emphasis here is on the fact that change has occurred, regardless of the larger historical perspective.

centralization of AI. This change is indeed innovative with regard to the behavior of Martha's Vineyard speakers, as evidenced by the age-grading of centralization (several times higher among 31-45 year olds than the over 75 group). However, a historical linguist might well discard the change in AI as retentive, which would force us to view the behavior of AI and AU as two separate and distinct phenomena rather than similar cases of centralization that are used for the same types of acts of identity.

The Martha's Vineyard example also illustrates why we have included 'focusing' as one of the ways in which an alignment event occurs. The speech of Martha's Vineyard already contained several variants of the changing phonemes, so it isn't entirely accurate to call what occurred there 'diffusion' in the purely geographical sense. Instead, it was focusing, making a given linguistic variant standard within a community.

The behavior of Druze in the Levant might clarify how an instance of focusing acts as an alignment event. The Druze are an endogamous, persecuted religious group living in close proximity to other social groups in what is now northern Israel and Jordan, and southern Syria and Lebanon. Unlike many of their neighbors, the Druze retain a retentive [q] realization of the Q variable in their speech, while the innovative realization [ʔ] is used by many of their neighbors, and this appears to be a change in progress. The resistance to this change is not entirely unique to the Druze, and in the recent past, as the change [q] > [ʔ] spread, other groups of course would have been characterized by this retention. However, the Druze are the primary group resisting this change, and [q] has become "a rule-of-thumb sign, in this area, for recognizing a Druze" (Blanc, 1953). That is to say, members of this community, which are in contact with the 'sociolinguistic marketplace' of possible realizations of the Q variable continue to focus their use of the [q] realization.

There is therefore a causal relationship in that being Druze makes one use the [q], data which is of incredible importance to understanding the historical linguistic development of this region. Of course, without knowing about the non-linguistic community boundaries, it would be impossible to separate Druze's use of [q] from that of say a remote Christian community which has also not shifted to [ʔ] due to insufficient network links, and this is one reason why retentions should be subject to a greater standard of evidence than innovations, but should not necessarily be eliminated entirely from consideration.

In terms of diffusion, there are many cases of what historically are retentions spreading through groups who had previously innovated replacements for those retentions. Fagyal et al. (2010, p. 2074) give a number of examples of this phenomenon, from the micro-level revitalization of forms in teenage in-group speech (Stenström, G. Andersen, and Hasund, 2002) to contra-hierarchical diffusion of archaic vowel realizations in Oklahoma and North Carolina (G. Bailey, T. Wickle, and Sand, 1993). They also develop a computer model which suggests that forms (whether retentions or innovations from a historical perspective) may fall out of favor only to be resurrected, meaning that what has become a retention can spread again through a speech community.

Another illustrative situation is that of Muslim Baghdadi (MB) Arabic. It seems quite likely that prior to about 1200 CE, it was a very similar to Jewish (JB) and Christian Baghdadi (CB) Arabic, but the fall of the Abbasids, the sack of Baghdad, and the growing importance of Muslim Bedouins in the immediately surrounding areas lead to the dialect becoming slowly closer to the nearby Bedouin dialects (Palva, 2009). MB Arabic now has a number of features which align it with nearby Muslim Bedouin dialects, but that is not to say that it shares a common pre-history with those dialects. That is to say that, for example, the MB use of the

[g] realization of the Q variable is not indicative of Hock’s “period of exclusively shared prehistory, during which [dialects] are in close contact with one another, but not with the rest of the [family],” but rather of a much, much later period. There are two alignment events (at least) which occurred here, in different places and at different times.³²

At the same time, innovations are simply more likely to be sociolinguistically relevant, and are much easier for the historical linguist to diagnose, since they will occur in a smaller subset of the data and display clear patterns of spread. Using retentive features, while theoretically possible, necessarily requires a higher standard of evidence. In the case of the Druze, we would need sufficient sociolinguistic data to understand how the failure to converge to the /ʔ/ realization of Q relates to their speech community boundaries. In the case of MB Arabic, a complete convergence to nearby Bedouin dialects would leave no indication that these dialects did not participate in the same prehistoric alignment events as those dialects; only because we have features which have survived that influence can we argue for a different origin for that dialect.

1.3.2 Diffusion within closely related lects

This section explores how the linguistic behavior changes over time, in-situ, within regional and similar scale speech communities, in the absence of major social changes (in the words of Ross, 1997, “speech community events”). Though

³² A historical linguist like Hetzron (1976) might argue that this is exactly why phonology is not necessarily indicative of genetic relatedness, and that we instead should use morphological features as diagnostics, but throughout my data I find numerous indications of near wholesale borrowing of demonstrative paradigms, and the spread of the Arabic definite article is a well established ‘areal’ feature which moved by some form of borrowing. The solution isn’t simply to winnow away all the least likely candidates for borrowing, but rather to try to take a more fine-grained view of diffusion and change.

there may be no specific changes to speech communities during the emergence and spread of linguistic forms throughout a region, the framework here suggests that any changes will be socially relevant in some way during their diffusion, though the social symbolism may differ between different social groups. This section will begin with a brief overview of social diffusion, i.e. the diffusion of a feature to encompass increasingly large speech communities, and then will discuss the more useful diagnostics of geographical diffusion.

1.3.2.1 Social diffusion

Social diffusion of changes is the process by which changes move through increasingly larger speech communities located in essentially the same geographical space. This diffusion can be diagnostic of speech communities, since they are the minimum unit which changes, and so a particular social class or group can be distinguished by the extent of a alignment event in a historical perspective.³³

As these changes diffuse between speech communities, some communities may not be able to or may be unwilling to integrate into the linguistic behavior of a speech community undergoing an alignment event, creating what one might call a SOCIAL ENCLAVE, similar to the notion of a geographical enclave found in dialect geography. Some researchers, following Blanc (1964), refer to social enclaves as ‘communal dialects,’ a term which also emphasizes the notion of community and shared identity. Social enclaves can be of various sorts. They can be gendered,

³³ Labov distinguishes between ‘change from below’, that is changes originating below the social consciousness of speakers, versus ‘change from above’ where speakers consciously adopt a change. However the level of social awareness seems to be relatively immaterial in terms of using changes as a diagnostic of social groupings, since changes diffuse along the lines of social networks and therefore at any given time even an ‘indicator’ is indicative of a particular social group even if it remains below the level of awareness (see discussion of this above in Section 1.2.1.2).

such that significant variation exists between men and women's speech in some places. Thomason (2007, p. 53) gives an example where men's speech was altered by their ritual register, while women, who didn't participate in these rituals, did not change, creating a gendered linguistic split.

Similarly, per Blanc (1964), the boundaries of a social enclave may be essentially co-terminous with the boundaries of a religious group, as they were in his research in Baghdad (see also Palva, 2009). This was the case with the Druze use of /q/ for the Q variable. The boundaries may be ethnic as well, with members of minority ethnic groups not participating in the sound changes of the majority speech community either as a result of segregation, or as a result of an unwillingness to align with the majority speech community, as discussed above with regards to African-Americans intergrated into white social networks, but maintaining AAVE features (Henderson, 1996).

These boundaries are not absolute of course, and speakers may style shift in different communicative environments, or may show different behavior if they belong to overlapping social groups than if they are in the core of a minority social group. For example, Labov (2011, chapter 16) notes in a small scale study of the use or lack of use of -s verb agreement, those members of the black community that have the most contacts and interaction with the white community use -s agreement more frequently than the core members of the black community, whose lack of -s agreement is close to categorical.

At the macro-level, with two regional speech communities increasingly differentiating themselves, we might expect sharper divisions between these communities to occur directly at the borders between them. However, this is typically not the case, with a more typical pattern being a continuum of dialect variation. This is due to the many overlapping allegiance which speakers have, with local

allegiances often stronger than national or other political allegiances.

There are cases where there is a sharper ideological and physical meaningful border that does create a clear isogloss line. As discussed above, the borders between modern nation states do create such sharp borders, as shown in Figure 1.2. There is the possibility that the same would be true of speakers on the borders of pre-modern empires, such as the border between the Arabs and the Byzantines, or possibly between the Fatimid and Abbasid caliphates. Even when there are firm borders, however, we have speakers who possess multiple identities which they indicate linguistically in ways that can vary situationally (but which relate to the identity they wish to index) as in the situation at the English-Scottish border described by Llamas (2010); Llamas, Watt, and Johnson (2009); Watt, Llamas, and Johnson (2010). Nonetheless, there are still correlations between speakers' primary identity and their use of language related to the emblematic 'center' for that identity (Llamas, 2007, pp. 600-602).

This suggests that diagnostically at least, a center-periphery model still holds for speech communities on the boundary between influential centers. The alignment events that differentiate those larger groups may arise in the center, but those on the borders are likely to be members of several different speech communities, and thus less likely to sharply contrast themselves against those on the other side of the border with whom they share speech community membership. However the degree of convergence to the center is related to the degree to which someone in these outlying speech communities feels a part of the community of the center, and we can use this as a diagnostic to reconstruct how centripetal or centrifugal these speech communities were.

1.3.2.2 Geographical Diffusion

The primary pattern of diffusion discussed in the sociolinguistic literature is diffusion along the urban hierarchy, with large population centers at the top of the hierarchy and smaller centers (or perhaps diffuse areas) at the bottom. This is contrasted with the more traditional wave model of diffusion, where innovations diffuse outward from population centers to the nearest adjacent locations spatially, but as Britain (2008, p. 623) notes, however, that evidence for wave-like diffusion has been rather limited. Instead most of the data shows that hierarchical diffusion, primarily from the top of the hierarchy downward, is a more common direction of diffusion (Britain, 2008, gives a number of examples of ‘hierarchical diffusion’ on pp. 623–626; Tældeman, 1995).

At the same time, there is a large quantity of evidence that innovations can move in the other direction, from rural locations to urban centers, but this diffusion still appears to operate along the urban hierarchy (Britain, 2008, pp. 627–628; Thomas Wikle and Guy Bailey, 1996; Thomas Wikle, Guy Bailey, et al., 1993). From a sociolinguistic perspective, this is an intriguing result, but from our perspective it simply suggests a general principle, that we might term the PRINCIPLE OF URBAN-RURAL DIFFERENTIATION: That highly populated and sparsely populated areas tend to diverge linguistically, reflecting the different speech communities involved in these alignment events and the differences in how they interact with other areas, as well as their attitudes towards one another.³⁴

³⁴ Tældeman (1995, pp. 273–274) notes that cities become more insular partially because they reject innovations from surrounding areas, resisting perhaps the less urban connotations of these innovations, i.e. urbanites do not want to be perceived as ‘country bumpkins’. So while there is a difference in social network connections that causes hierarchical diffusion, as discussed in 1.2.1.1, there is also a difference in allegiance and identity also. See also the discussion of the loss of feminine plural marking in Iraq in Chapter 6.3.3.

Social networks in urban areas also tend to contain more weak links, as speakers have relatively more frequent and intensive contact with relative strangers than they would in a rural environment, a situation that leads to faster absolute rate of change in cities, which coupled with more frequent interactions with urbanites from other locations leads to faster diffusion of changes through the urban speech community as opposed to rural communities (L. Milroy, 1980; Tældeman, 1995, pp. 276–277).

In the Arabic speaking world there is a tripartite rather than a bipartite division — while there is an expected urban-rural dialect differentiation in many areas (particularly the Levant), there is a third differentiation between sedentary dialects and those of nomadic Bedouin. The general differentiation that occurs between urban and rural dialects is expected, as per the research discussed above.

The division between sedentary and Bedouin dialects is due to similar factors. Bedouin social networks are often even more dense than rural networks, and in modern Bedouin groups (and likely in pre-modern Bedouin society) marriage is largely endogamous, with marriage to non-Bedouin disfavored (Henkin, 2010, Chapter 2). Though some Bedouin groups have more frequent interactions with rural groups (but rarely with large urban centers, as a result of their subsistence patterns that require large, open spaces), these interactions are much less frequent than urban-rural interactions, and certainly less than urban-urban interactions.

It is not just different social network connections at work, but also social allegiance which plays a role in differentiating sedentary and Bedouin speakers. For example, modern Negev Bedouin look down upon sedentary farmers (*ibid.*, Chapter 2), and evidence from early Islamic sources suggest that urban speakers

likewise looked down on Bedouin and the Bedouin lifestyle (Bashear, 1997)³⁵ Moreover, since Bedouin are extremely mobile, they can move into an area where they have not previously resided, and therefore where they have not participated in the local linguistic alignment events, creating an illusion of a greater linguistic difference between Bedouin and sedentary populations (see Section 2.5 for more on this).

Another approach to understanding geographical diffusion of changes is the attempt to link specific geographic patterns of diffusion with likely historical scenarios, as is common in the German and Dutch dialect geography tradition. However, it seems that this literature is not significantly more illuminating than simply understanding the principles that underlie this diffusion. The basic principle behind this approach, Behnstedt and Woidich (2005, p. 147) quote Goossens (1977, p. 81) as stating:

Diese Formen sind nicht das Ergebnis des Zufalls; sie sind im Gegenteil durch *spezifische Sprachbewegungen* [their emphasis] zustande gekommen. Man kann deshalb versuchen, aus der Form der Areale die Richtung der sprachgeographischen Verschiebungen abzuleiten.

These [cartographic] forms are not the result of chance; they are rather the result of *specific linguistic movements*³⁶ One can, therefore, attempt

³⁵ This anti-Bedouin sentiment was later reversed, or possibly overlaid, with a romanticization of Bedouin life in later periods, as they began to be seen as linguistically ‘pure’ (and well might have been more conservative linguistically, as a result of the tight social networks of Bedouin social groups) among other qualities. This is all tied to a largely nation-building/imagining movement, a topic of great interest but which is well outside the scope of this dissertation. In any case, the actual linguistic impact of this (apparent) change in attitudes is not clear.

³⁶ “Sprachbewegung” here is used in a sense of linguistic diffusion rather than the movement of speakers, i.e. “Sprecherbewegung”.

to derive the direction of linguistic-geographic movements from the form of the [geographic] areas.

Weijnen (1977) has, on the other hand, expressed strong reservations about the assumption that similarity of geographical configuration is a strong indicator of similarity of historical events. Many map configurations can be read in multiple ways, either as an example of the diffusion of innovations or as the maintenance of ‘relic’ forms. He even quotes Goossens (1977) as saying, “the form of the dialect area in [and of] itself is no evidence of the direction of the derivation. This argument can only be used to support the other arguments (his translation)” so that as Weijnen states, “historical testimonies and [...] factual history are just as important as the geolinguistic picture”, to which I will add here the testimony of sociolinguistics and the analysis of speech communities. Therefore, changes which present a distinctive geographical pattern will be evaluated on a case-by-case basis, and we will not further explore the mapping between geographical configurations and historical circumstances here.

1.3.3 Diffusion between distantly related lects

An advantage of the speech community approach is that it allows us to model multi-lingual speech communities, that is groups connected by social networks that view themselves as part of one larger linguistic community, in spite of the differences in the individual lects involved. It is not sufficient for two speech communities speaking different lects to be adjacent — for any significant linguistic interaction to take place, the communities must have connecting social networks and some sense that their languages are commensurate and that the members of the speech community view the languages involved as available targets for linguistic production. If two geographically adjacent communities don’t have either

the network links or a sense of allegiance (see section 1.2.2.1 regarding speaker attitudes and mutual intelligibility per Wolff, 1959), it is unlikely that there will be significant linguistic results.

This section will discuss the diagnostics of long-term multilingualism, as opposed to the results of a loss of language by one of the participating speech communities (i.e. shift) which is discussed in section 1.3.5. As discussed in that section, the results of shifts are difficult to detect if there is no fairly lengthy intervening period of multilingualism, the effects of which are clearer and better studied in the linguistic literature. Almost all of the examples of slight to heavy substratum interference in Thomason and Kaufman (1988, pp. 121-146) are the result of a lengthy history of multilingualism and contact between members of speech communities using different languages, and even when that period was extensive (ca. 300 years for Frankish-Gallo-Romance contact, 200 years for Norman French-English contact) actual structural changes as a result of ‘substrate influence’ may be minimal. That is to say, the fact that a group ‘shifted’ to using another lect is not what causes significant, diagnostic changes (except in some areas) — it is the period between the initial contact of the speech communities and the disappearance of speakers of the shifting lect that has the largest effects on language.

This is particularly important for the history of Arabic because we have fairly detailed knowledge of the underlying languages which shifted to Arabic, but which undoubtedly existed in a stable bilingual relationship with Arabic for at least several centuries before shifting — in the case of North Africa, some degree of Arabic-Tamazight (Berber) bilingualism has existed in places for nearly a millenium and a half. Moreover, since the definition of what constitutes another dialect or language is not determined by a linguist, but by speakers, the effects of multi-lectalism such as calqueing of forms (instead of diffusion) can be seen even

between what are in modern terms, different ‘dialects’ of Arabic. For example, as shown in Section 5.3.2.1, and elaborated in Section 6.3, the *ḍawla*-type demonstrative forms are actually calqued based on *hawla*-type demonstratives.

Broadly speaking, we can reconstruct three major aspects of speech community interaction from the linguistic results of these interactions: 1) How speakers viewed themselves in relation to other speech communities, 2) The relative dominance of these communities, 3) The degree to which they were integrated into one another and 4) The degree of multilingualism that was present in the communities.

1.3.3.1 Degree of difference

Since it is speaker’s conceptions of what makes languages different, within certain bounds, the results of diffusion between ‘different’ lects can be diagnostic of how speech communities viewed one another, though this is primarily true for closely related languages. If two lects are almost entirely identical, then the spread of changes will be along the lines of normal diffusion, and the new forms spreading through the speech communities will simply replace the old forms completely.

However, if there has been some divergence between the lects, but “speakers perceive the two lects as variants of each other, then they may borrow bound morphemes from a paradigm in the secondary lect into the corresponding paradigm of the primary lect (Ross, 2005, p. 190).” This is not simply true of bound morphemes – any morpheme could potentially be moved between the two lects, however the most likely candidates for borrowing would be high frequency morphemes, or the highest frequency members of a paradigm. Lower frequency members of the paradigm are less likely to be available and salient to the speakers of the recipient lect, and so this process would produce mixed paradigms, where the highest

frequency members of that paradigm are borrowed, but the lower frequency members reflect the original forms in that lect.

Morpheme borrowing is also diagnostic of very permeable speech community boundaries, either with the recipient speech community reorienting itself as a member of the donor speech community, and so replacing its linguistic material with that of the donor.

In contrast, speech communities with identical ‘linguistic distance’ might instead copy the forms of the other community through the process of metatypy (i.e. calqueing). This would be diagnostic of an attitudinal stance that views the two speech communities as integrated into a larger community, enough to realign their linguistic systems towards one another, but not close enough to directly borrow that material (Ross, 2005, see further).

1.3.3.2 Dominance

The notion of (sociolinguistic) dominance here is a conflation of a number of factors, and no matter how granular a definition we devise, it will still remain a simplification on some levels. Dominance in a general sense is related to the sociolinguistic place of a language in society, i.e. whether a language is a major language of scholarship, communication, or religion (English, Latin, Sanskrit, Arabic at various time periods), whether a speech community is numerically larger than another, and whether that group is politically dominant or not. These are not all related variables, and a more thorough treatment would separate them.

Linguistic dominance can also be broken down into a simpler bipartite division: The first part is the likelihood that speakers will acquire and use another lect (and which group will be the one acquiring, such that a subordinate group is more likely to be multi-lectal while a dominant group is less so), and the second

is the likelihood that they will adopt features from this other lect (whether as a result of positive social valuation or simply as a result of being more wide-spread statistically and thus more available to learners.)

One of the primary diagnostics of dominance is the degree to which one language appears to influence another, with a subordinate speech community tending to be bilingual with the dominant community remaining monolingual (Matras, 2009, p. 59). The practical result of this is that the subordinate speech community tends to acquire a larger quantity of loanwords from the dominant speech community, and tends to undergo more structural convergence towards the dominant speech community's lect than vice-versa.

However, it may be better to look at the flow of loanwords through the lens of 'expertise' rather than dominance. The domain of expertise of a given speech community tends to be in their language, and so members of other speech communities who wish to work in those domains tends to borrow that lect's vocabulary. Hence, it is not unusual at all for conquering speech communities to borrow vocabulary for flora and fauna from the local population. Similarly minority speech communities which have distinct ethnic cuisines tend to contribute the vocabulary for those foods to the dominant speech community (e.g. bagel, knish). For religious communities, the language associated with religion may provides loanwords into their native language. These are generally borrowings where the recipient language has no equivalent, and therefore they must either borrow a word, or invent their own.

However, the dominant language also tends to become the language of education, trade and technology, so that the terms for these important areas of expertise tend to be borrowed, even when there is already a local equivalent. On the whole, we expect more loanwords to move from dominant to less dominant

speech communities. This also means that borrowed terms which *are* in the expertise of the subordinate groups are particularly strong diagnostics of subordinate status.

There may also be something of a dominance see-saw, especially at a lower level of analysis, between the level of small-scale and regional speech communities.³⁷ As a new group becomes dominant in an area (through conquest, migration, changing circumstances) they are still less numerically and culturally dominant than their neighbors. At this point, the ‘dominant’ language (from a historical perspective) may borrow local terminology and even have a certain degree of multilingualism that the subordinate speech community does not have. However, as the community grows in power and numbers, influence moves in the other direction, with a likely balance point of stable, relatively equal multilingualism occurring for some period of time. Finally, in a shift situation, as there is increasing linguistic attrition in the historically subordinate community, the influence of the dominant community will be greater, and the subordinate community less until the subordinate speech community is subsumed.

On the other hand, mutual influence by the lects involved is indicative of relative equal dominance, where no speech community is clearly dominant over the others and there is a high degree degree of mutual multilingualism.³⁸ Such situations often appear to have some sort of taboo or resistance to using foreign lexemes, so that it is largely structural convergence that results from these situations (Matras, 2009, p. 60).³⁹ Examples of such situations include the Amazonian

³⁷ See also the account of language attrition in Section 1.3.5.

³⁸ Without multilingualism, equal dominance is probably not possible to diagnose from the historical evidence.

³⁹ Indeed, it would be rather difficult to diagnose a situation where two equally dominant lects heavily influence each other lexically and structurally.

sedentary agriculturalists system of linguistic exogamy where lexeme borrowing and code-switching are taboo, but strong structural convergence has occurred between the participating languages (Aikhenvald, 2003), and the convergence in Kupwar languages (Gumperz and Wilson, 1971). In the Kupwar situation, with no explicit taboo on borrowing words, there is some content borrowing and some inflectional morphemes are borrowed as well, but there is very little borrowing of paradigmatic inflectional morphology, which “seem to be at the core of the native speakers perception of what constitute ‘different languages’ (*ibid.*, p. 162)”.

Note however that even in the Kupwar situation, subtle differences in language dominance are actually reflected by how far lects converge towards one another, with speakers converging towards Marathi more than towards other lects, reflecting its status as a relatively dominant language of the wider areal speech community and its lack of social marking within the village. Thus, the direction of convergence is an good diagnostic of sociolinguistic dominance, which can then be related to speech community relations and notions of allegiance.

1.3.3.3 Integration

Just as with diffusion between closely related lects, speech communities may be more or less integrated into a nearby speech community that uses a more distantly related lect.⁴⁰ If we have a situation of shift, we can assume that the shifting speech community was relatively less dominant than, and was reasonably well integrated into, the speech community which they shift to. On the other hand, if a community resists shifting for a long period of time, i.e. they form an enclave, we can diagnose that there were certain attitudinal or network forces which helped

⁴⁰ See Schumann (1978, 1986), for this in the context of language shift which occurs within a small period of time (one to three generations).

them preserve their original lect. Even if they do shift, they may maintain certain features which continue to distinguish them from the larger community, as is the case with Hispanic or AAV English, which would be diagnostic of an extreme discrepancy in dominance between the communities, with a concomitant strength in the internal cohesion of the non-dominant speech community.

1.3.3.4 Multilingualism

The degree to which members of contacting speech communities are multilingual also influences the course of contact, and therefore we can diagnose something about how proficient multilingual users of the lects available were within a given community, based on linguistic evidence.

The degree and spread of multilingualism dictates what kind of materials can transfer between lects. Words of material culture and regularly used concepts (including e.g. religious terminology) do not necessarily require a high degree of multilingualism, and can be borrowed relatively easily. On the other hand, borrowing of verbs, pronouns, and bound morphology generally require significantly greater familiarity with the source language for these materials to be borrowed, as well as a perception that the languages are comparable enough for this borrowing to occur. Not all of these grammatical forms are equally borrowable, and Matras, 2007, p. 61 gives the follow scale of how frequently different forms are borrowed:

Nouns, conjunctions > verbs > discourse markers > adjectives > interjections > adverbs > other particles, adpositions > numerals > pronouns > derivational affixes > inflectional affixes

Similarly, structural convergence is also a strong indication of high proficiency multilingualism. In general, the existence of multiple lects in the mind of a

single speaker appears to cause interference on both the native and acquired lect(s), leading to structural convergence. Ross (2007) calls this convergence METATYPY, defined as “a diachronic process whereby the morphosyntactic constructions of one of the languages of a bilingual speech community are restructured on the model of the constructions of the speaker’s other language.”

These morphosyntactic changes are likely to be derived from changes on the level of semantics or pragmatics. Silva-Corvalán (2008) argues, based on examples of apparently syntactic interference between Spanish and other languages, that syntax itself is nontransferable between languages. The appearance of features which share syntactic similarities (but generally not phonetic form) is evidence, she argues, that:

[...] what is transferred is *not* a syntactic structure, but the semantics or the pragmatics of a construction, which is then linked to a closely structural parallel in the recipient language. (p. 221)

An example of this process is the metatypic spread of evidentiality marking. As Aikhenvald (2004, §9.2) notes, evidentiality is typologically quite rare, but is frequently clustered in a small geographical area, suggesting that it diffuses quite easily. The cause seems to be that multilinguals in one lect have no concept of the need to mark evidentiality, while in the other they do. As multilinguals become aware of this pragmatic requirement in the other language, they transfer the same pragmatic notion to their own language (or rather, they more likely have an integrated system in their mind which now demands fulfillment of that pragmatic function) and use material from the language which lacks formal evidentiality marking to mark the same concept. Eventually, this grammaticalizes to create specific grammatical markings.

Languages can also lose semantic-pragmatic categories as a result of metatypic mental convergence. Again, an example of this is the *loss* of evidentiality marking as a result of shifting dominance patterns. Adamou ([forthcoming](#)) reports on a situation of trilingualism in the Pomok-speaking region of Greece, quite near to Turkey. Pomok, a dialect of Bulgarian, has traditionally had a system of evidentiality. However, a change appears to be in progress that is causing the loss of evidentiality. Though Adamou ([ibid.](#)) presents a multiple causation analysis, one major cause appears to be that Greek, which lacks evidentiality, is becoming increasingly dominant, at the expense of Turkish. Most people still speak all three languages, but Greek is more dominant, and this seems to be influencing the categories in Pomok. This is also an excellent illustration of the parameter of dominance — as Adamou ([ibid.](#), p. 22) notes, “For Pomok, it is not migration that determined this change, but a change in the hierarchy of the contact languages.”

Another diagnostic of which direction metatypy is moving in can be determined by the nature of the replicated material. The grammaticalization process following the transfer of categories takes time — for example, the Hup evidentiality morphemes in Epps ([2005](#)) are apparently newer and more transparent than those in the other Amazonian languages which are clearly older. This gives us an idea of the dominance relations between the two languages, as the more dominant languages’ pattern seems to be more likely to be replicated (i.e. it is the model for what a language is) versus the subordinate speech community’s language.

We can, therefore, diagnose some of the internal divisions and relations between speech communities who speak disparate languages but which are bound together as a single, larger community by looking at how those languages have been affected by these interactions. The continuum of borrowing from content

words to functional morphemes reflects, in a gross sense, the degree of multilingualism which existed in the larger speech community. Metatypic change requires relatively widespread, high proficiency multilingualism, and so structural convergence is diagnostic of this, while the direction in which convergence occurs may indicate something of the degree of dominance. Dominance also tends to determine who is multilingual, and therefore how linguistic material flows, while the degree of integration determines whether a group participates in a larger multilingual speech community.

1.3.4 Newly established communities

To understand how Arabic spread and changed over time, it is vital to understand how newly established communities choose and focus a given variety of language. The Islamic conquests resulted in the creation of huge cities, largely established *ex nihilo*, and we can leverage modern sociolinguistic research to extract as much information about the varieties languages as possible. Moreover, newly established communities have a linguistic robustness which may explain the success of Arabic despite the much lower numbers of conquering Arabs relative to the conquered people speaking other languages (see Section 3.2.2.1).

When speakers settle together in a new area, especially if that area lacks a strong speech community, they form a linguistic beachhead, one which has an outsized impact on the linguistic behavior of other speech communities that enter the area. The reason for this outsize impact is the principle of FIRST EFFECTIVE SETTLEMENT, a term taken from cultural geography. Zelinsky (1992, p. 13), the originator of the principle, defines it as follows:

Whenever an empty territory undergoes settlement or an earlier population is dislodged by invaders, the specific characteristics of the first

group able to effect a viable, self-perpetuating society are of crucial significance for the later social and cultural geography of the area, no matter how tiny the initial band of settlers may have been.

The cultural patterns, including the linguistic ones, of the first settlers therefore appear to set the basis for the language and specific linguistic features of a given place, even if speakers of other languages and dialects move into the area. Note, however, that this is a tendency and not absolute. As Labov (2001, p. 504) states:

In any one generation, if the numbers of immigrants rise to an order of magnitude greater than the extant population, the doctrine may be overthrown, with quantitative changes in the general speech pattern.

Zelinsky (1992, p. 20) gives the example of the small Scandinavian settlements along the Delaware Valley which were “promptly swamped by subsequent British and Teutonic colonization.” Even though some enclaves did manage to survive, the general principle is quite strong.⁴¹

⁴¹ Though the principle of first effective settlement is essentially an observation of a historical trend, and therefore can be taken ‘as is’ for understanding historical events, Croft’s (2000) notion of “convention” may be the mechanism that underlies it. He discusses convention as the way in which communication is made possible at all, a so-called “coordinating” device that allows speakers to be on the same page. For this to work, a convention must be “a regularity in behavior that almost everyone on the community conforms to, almost everyone expects almost everyone else to conform to, and almost everyone would prefer any new member of the community to conform to. (p. 97)” It is the last clause here that underlies the first effective settlement principle — if a community expects all new members to conform to their established pattern in order to be a part of their speech community, then that kind of conformity must exert strong pressure on immigrants. On the other hand, if the immigrating speech community is numerically large enough and prestigious enough, they can force the flow of convention in the other direction, expecting the original inhabitants to accommodate to their notion of convention (as happens in some shift situations).

1.3.4.1 Dialect Mixing to Focusing

In many cases, these settlers come from different speech communities, and in their new community they must negotiate the linguistic code for their new speech community. In certain cases, of course, some of the speech communities involved are likely to maintain their lect and to try not to participate in the changes in the rest of the speech community, but unless there is incredibly strong internal group cohesion, changes will probably diffuse to their speech community. The process of reaching a linguistic compromise for the lect of the newly established speech community is often referred to as KOINEIZATION. This term is rather polysemous in the linguistics literature (Siegel, 1985), but will be used here to refer to the formation of a hybrid lect as the result of multiple speech communities migrating to a new, tabula rasa settlement.⁴² The goal of this section is to look at how we can diagnose whether a speech community in the present is the result of the integration of several speech communities in the past, and whether we can tease out some of the linguistic behavior and extra-linguistic circumstances that surrounded that integration.

In the framework of Trudgill and Kerswill (2005, p. 200), there are three stages that occur in the koineization process. In the first stage, the speech communities which have been brought together in the new settlement remain stably separate, but linguistic forms which are symbolic of membership in a stigmatized community tend to be lost or suppressed in intergroup interaction.

When the speech communities begin to integrate, a period of extreme variability in linguistic behavior seems to result, with some leveling to forms common to both groups but typically a bewildering level of variability which exceeds that

⁴² The inputs to koineization are what would be called ‘dialects’ in so far as that is a meaningful category. If the input languages are too different, the process and result would likely be different.

of the Stage One group's individual linguistic behavior. The forms used by the speakers at this stage seem to be in the process of being suggested as candidates for the symbolic forms that will be emblematic of the new speech community that is being formed. Must surprisingly, these forms are not necessarily the result of inheritance from parents, but reflect a "shopping-basket" effect (Trudgill and Kerwill, 2005, p. 210) where the totality of the variation present in the community is available for recruitment to symbolize groups.

As this variation is eventually selected by different groups as being symbolic of their speech communities (i.e. reallocation of geographical variation to sociolinguistic markers), the norms eventually focus and a new speech community emerges, based in a shared allegiance between members of the speech communities in the settlement. Relative to the initial and second stage, the linguistic behavior is much more unified, and this unity probably allows for the implementation of conventionalization that would encourage in-migrating groups to assimilate to this new norm.

There are four major linguistic effects of the integration of these speech communities: 1) MIXING, i.e. the preservation side-by-side of forms found in the speech of all the communities involved, 2) LEVELING, the reduction of variation between the lects by choosing a form from only one lect, 3) INTERDIALECT FORMS, the creation of intermediate forms between the forms found in the participating lects, and 4) REALLOCATION, the repurposing of variation between the lects for other uses (sociolinguistic markers, stylistic markers, even occasionally allophonic variation).

Mixing can be helpful for determining what lects were used by the original speech communities. The forms in these lects which are from clearly different historical developments might be preserved side-by-side, and so show a fairly clear

picture of which groups participated in koineization.

Levelling in koineization is somewhat controversial, particularly with regards to what exactly gets leveled and why. Siegel (1985, p. 364) quotes Dillard (1972) as defining dialect leveling as “the process of eliminating prominent stereotypable features of differences between dialects”. Though the framework here recognizes all linguistic variation as at least (historically) having represented or still representing a symbol of group difference, at a given time and place some variation will be more significant and more socially marked, and it is this difference that tends to be removed in the first stages of koineization. These forms are often minority forms, or forms associated with minority populations. It is also a question of which group finds a feature stigmatized and stereotyped, and whether the linguistic dominance of that group is sufficient to cause the other to cease using certain forms.

Trudgill (2004) argues that the forms which survive koineization are those most frequent forms in the community, regardless of social valuation (see also Trudgill and Kerswill, 2005, pp. 202–203; Siegel, 1987; Mesthrie, 1992).⁴³ However, this theory is based on work on the totally-new dialect creation in New Zealand, and he explicitly excludes non-tabula rasa new city situations (like Milton Keynes, Kerswill and Williams, 2000) from his theory, since they are still connected to the networks of sociolinguistic values of their origins, and therefore could still participate in alignment events with their source population. Kerswill (2010) argues

⁴³ Note that it is not the relative population size of speakers so much as the population size of those speakers who have a given feature. Though the two are generally likely to be close to congruent, this is not always the case, so that a plurality that uses one form could be outweighed by a majority formed by different groups using a different form. This may simply reflect to the number of examples of each variant in the speech of the community, such that children learn the most frequently heard variants, as proposed by statistical learning approaches to language acquisition (Bybee, 2001).

strongly against this view even in apparent tabula rasa situations, since, as he notes “a tabula rasa quickly ceases to be a clean slate” as the community accumulates social valuations for the forms that are used, even if those social valuations are not the same as those in their homelands.⁴⁴

Diagnostically then, if we find that the speech community with a new-city history lacks certain features which appear to be marginal in another dialect area, or which we knew were dispreferred at a certain time period, this may indicate that they were eliminated due to the koineization processes. Alternately, this may reflect different population sizes of the original population, and the further the new speech community is from its parent speech communities, the more likely it is that demographics are more deterministic than social valuation. To determine which of these possibilities is more likely, we would need additional historical and non-linguistic testimony from our sources.

Reallocation is the re-assignment of a variation that marked one axis of variation (geography, class, etc.) to mark a different axis of variation (gender, neighborhood, etc.). Quite frequently, this variation gets reused to mark different speech community boundaries, what Britain and Trudgill (2005) refer to as ‘socio-stylistic reallocation’. For example, in Amman, Jordan, the use of [g] vs. [ʔ] realizations of the Q variable was reallocated from a marker of geographical origin to a complex marker of gender and social group (Al-Wer, 2007). Socio-stylistic reallocation is complex, and often requires a detailed understanding of the history of the speech communities involved. It can be useful as a diagnostic, though there is no single straightforward interpretation.⁴⁵

⁴⁴ For more on this debate, see Trudgill (2008) and the responses to this in volume 37 of *Language in Society*.

⁴⁵ See Palva (2009) for an admirable use of reallocation in understanding a historical series of events.

Interdialect forms appear to be a sort of compromise solution, where speakers choose an in-between form that splits the difference between their dialects. This is probably a result of forms where the social valuation of both forms is equalled valued or dispreferred, and which, like mixed languages, represents a new community marker which reflects those shared origins. Mixed forms can either be an intermediate phonetic realization, especially for vowels which vary continuously, a blend of two morphemes, or various insertions as a result of hypercorrection (for numerous examples see Trudgill, 1986, pp. 62-78). Interdialect forms, therefore, can be diagnostic of which forms were approximately equally valued in the contact situation.

Finally, many studies in koinization see a role for a general process of ‘simplification,’ a notoriously vague and therefore methodologically difficult concept. Trudgill (*ibid.*, pp. 102-7) acknowledges the danger of using such a blunt notion, and his attempts at a more focused definition themselves becomes increasingly large and vague, calling on notions of paradigmatic regularity, transparency of form-function relations, and the yet more nebulous notion of markedness (for criticisms of markedness as a category, see Haspelmath, 2006). Such a definition is extremely blunt and ripe for abuse, and has been used in ways which seem suspect at best to this researcher (see for example Penny, 2004). We will not, therefore, call upon simplification in order to understanding speech community integration.⁴⁶

⁴⁶ See also McWhorter (2007), who argues for adult language learning as a ‘simplifying’ factor in the development of various world languages, including Arabic, but whose framework largely relies on a notion of grammatical simplicity. As Riddle (2008) notes, low grammatical complexity often leads to an increase in lexical complexity. From a SLA perspective, a bounded number of grammatical rules seems significantly easier to acquire than a great deal of open-ended lexical complexity. Moreover, L2 learners tend to do better at grammar and morphology than they do at phonology, and so a measure of phonological simplification would better indicate adult group learning events. Note that for Arabic, such a measure would *not* show significant simplification as the quite large phonemic inventories of Arabic are largely intact, with the exception of the

1.3.5 Shift to another lect

The process by which one speech communities (over a period of time) abandons their language and acquires that of the other speech community, that is, SHIFT, leaves diagnosable traces in the linguistic-historic record. Shift happens when two speech communities, one dominant over the other, are brought together.

1.3.5.1 Direction of shift

Shift is likely to occur when a group moves into the territory of another and adopts their language (hence the first effective settlement principle). On the other hand, in-migrating groups do not always shift their language — they may maintain it, or they may acquire the language of their new community but retain certain features that mark them off as a separate group. In rare cases, they may be a large enough and sufficiently dominant group to cause the already extant populations to change towards the language of the new migrants. The question is then, why do different example of migration lead to different outcomes?

The primary factor is population density, related to Labov's proviso that a supplanting community must be significantly more numerous to obviate the principle of first effective settlement. Ostler (2005), attempting to understand in a broad, historical view why some languages succeeded in supplanting others, argues that areas with relatively high population density were more likely to resist language change. For example, Egypt was able to maintain its languages in the face of invaders for several thousand years, until finally shifting to Arabic. On a smaller scale, Buchheit (1988) analyzes the factors that caused some German-speaking Mennonite communities to resist shifting to English longer than others.

merger of interdentals to coronal stops in many dialects, a feature that seems quite old.

Total percentage of German speakers relative to English speakers was one of the most significant variables for delaying the shift to English. The corollary of this is that significant *depopulation* through war or disease is also likely to make it easier for migrants to displace or absorb previous speech communities.

This also relates to settlement patterns. Moving into an already established city puts the first effective settlement principle into play, while founding an entirely new city creates a stable foundation population that forms a strong speech community. Those offering goods and services in the new city from neighboring areas would be more likely to have to converge to the language of the new city rather than using their own language. Even if a group settled in an already established city, deliberate segregation would help maintain the integrity of the speech community and its language.

The relative social importance of the languages involved also plays a role in what happens to the merged speech communities. If one of the languages is the key to upward mobility, it is more likely that speakers of other languages will become bilingual in that language, and may shift entirely. Ostler (2005) suggests that for Chinese, the examination system that required high proficiency in Chinese, and which offered a meritocratic path to upward social mobility, played an important role in maintaining the Chinese language and prompting in-migrating groups to acquire Chinese. Arabic may have played a similar role, with open access to Arabic language education in mosques allowing for social mobility, but largely within the Arabic language.⁴⁷

⁴⁷ Another factor suggested by Ostler (*ibid.*) is the learnability of different languages, a function of typological distance in his framework. In the very short term, within two or three generations, this may be a factor in language acquisition by adults, as a language with plentiful cognates and similar grammar may be easier to acquire than a more distantly related one. However, once child language acquisition begins to occur (if the communities are not too segregated), it is no longer

1.3.5.2 Linguistic outcomes of shift

There is a tendency to see the linguistic results of shift as somehow reflective of ‘incomplete group second language acquisition’ (see for example Winford, 2003, Chapter 7). Such an approach is based on the resemblance between some of the outcomes of individual second language acquisition (the positive and negative transfer of categories from the L1, simplification of certain aspects of the L2 interlanguage), and the results of large scale language shift. However, in many cases of in-migration to an area with another dominant lect, language shift occurs rapidly, in the space of three generations, with the third generation acquiring the language fluently as they integrate into the speech community at large (*ibid.*, pp. 236-7). Even in longer term shift situations, child learners likely have a great deal of access to input in the second language, and it seems unlikely that they would have quite the same difficulties as adult language learners.

In cases where we do seem to have significant substrate influence as a result of shift, it actually looks like the shift was preceded by long periods of high proficiency bilingualism. Many of the so-called features of substrate interference look more similar to the kinds of changes brought about by metatypy (see Section 1.3.3). That is to say, substratum influence isn’t largely the result of poor language acquisition, but rather the opposite, caused by high proficiency bilingualism on the part of the (eventually) shifting population.

Winford (2003, p. 258), based on Batibo (1992) even gives a three stage

terribly relevant. At the time scales discussed here this seems to be unlikely to be a barrier, and examples of rapid shift between typologically distant languages are numerous. Indo-European languages were quite capable of becoming dominant among Amerindians in both North and South America, and typological differences between the languages don’t prevent the creation of mixed languages with input from quite typologically divergent languages, such as such as Michif (Cree and French) or Media Lengua (Quechua and Spanish). Similar, Turkic languages were able to become dominant in Asia minor, in spite of the existence of in-situ Indo-European languages.

account of language shift leading to attrition and death: the first stage is monolingualism in the L1, then growing bilingualism with the L1, then the L2 becoming dominant, and finally loss of the L1 and shift to the L2. Both the second and final phases require not only bilingualism, but opportunities for child acquisition of the L2, suggesting again that it is not some inability to acquire the L2 but rather other factors that account for substratum influence.

On the other hand, shifting speakers may not actually have perfect acquisition of the target language as their goal. Winford (2003, p. 237) gives the example of minority groups such as Hispanic and AAVE speakers in the US which maintain distinct ethnolects of English that are otherwise native-like in production, rather than adopting mainstream English dialects, even though members of those communities have ample opportunity to acquire the mainstream dialect at a young age. Similarly, he also criticizes the notion of ‘fossilization’ from SLA, the idea that second language learners cease learning at a certain point, quoting Escure (1997, p. 275) as saying that it “fails to capture the dynamic, innovative and — at least subconsciously — intentional use of old features” to preserve the distinctiveness of the shifting groups identity. That is to say, the shifting group maintains certain features that are emblematic of their group identity, and what could be more emblematic than features which reflect their original language?

At the same time, we need to account for phenomena such as the loss of categories, especially phonological, in the acquired language, even when members of the community likely had access to that language in childhood. For example, Chadian Arabic lacks the emphatic and pharyngeal consonants typical of Arabic dialects, and has gained several other phonemes (including implosives) that all suggest a certain amount of influence from neighboring languages. One might propose that the initial group of learners lost the emphatics and this became em-

blematic of their language, that is to say that initial acquisition provides the raw material for the expression of identity through divergent realizations of the acquired language. However, it is also possible that like the spread of Khosian clicks to Bantu languages (Thomason and Kaufman, 1988, p. 133), the non-use of emphatics represents a general set of areal linguistic practices, and therefore would be the result of bilingualism and metatypy rather than failed acquisition per se.

Diagnostically, then, greater degrees of substrate influence reflect a longer period of high proficiency bilingualism, combined with a strong maintenance of group identity on the part of the shifting speakers over a long period of time. It does not reflect a lack of access to the language that is being shifted toward, or an inability of the shifters to acquire that language correctly. The degree of substrate influence also probably correlates with the strength of speech community allegiance and identity, so we can get a better idea of speech community relations by looking at substrate influence. If the entire population shares these features, rather than a specific ethnic group, we can also posit that the ethnic shifting group was large and dominant enough for their features to become the norm, and that the speech community boundaries between those groups at some point largely dissolved.

1.4 Summary

The basic project of this dissertation is to reconstruct the variation that existed in the Arabic language immediately preceding the Islamic conquests, and in this chapter I have explored the challenge of developing a framework to undertake that reconstruction. I have shown that the language-centric historical approach which aims at reconstructing a ‘proto-language’ is not entirely appropriate for this task.

Instead, I build on recent work in historical linguistics which instead focuses on ‘speech communities,’ normally defined in terms of their social network structures. I propose an additional criterion for defining a speech community, the notion of shared allegiance between members of a speech community. I demonstrated that this notion has a sound empirical foundation in sociolinguistics, and that the notion of allegiance makes it easier to integrate both linguistic and extralinguistic information into our reconstructions of speech communities in a principled way. Finally, I discussed the various types of contemporary linguistic evidence that can tell us about the relationships between speech communities that existed in the past, whether they speak similar or dissimilar languages.

This framework is intended to be a general one that would be applicable to languages worldwide which have a great deal of modern diversity, and where we need to know not only about their distant past, but about a particular time in their linguistic development. In the following chapters, and in Chapter 6 in particular, I will show how this approach can be used to reconstruct an incredibly amount of information about the pre-Islamic Arabic speech communities, building on the data from the preceding chapters.

Chapter 2

Pre-Islamic Arabia: Tribes and Social Organization

The purpose of this chapter is to question some of the long held notions about tribal organization and subsistence patterns among Arabic speakers in the pre–(and post–) Islamic world by synthesizing studies from anthropology and other fields. Very traditional treatments of Arabic speaking peoples, or peoples referred to as Arabs, have tended to focus on their role as nomadic pastoralists,¹ an image which itself was enhanced by a glorification of desert life following the Islamic conquests (Bashear, 1997).

This chapter asks what kinds of lifestyles were practiced by Arabs and similar people's in the pre-Islamic world, and what this means for the linguistic interactions between those groups as well as for modern dialect geography.² It also

¹ For example, talking about the spread of Arabs generally throughout history, Caskel (*EI2*, “*al-Arab (ii)*”) “the expansion consists usually in the emigration of large or small nomadic groups, rarely in that of groups with permanent habitations.” Note the emphasis on nomadism, and the downplaying of any role for settled people.

² The linguistic import of social organization is a topic which is a growing area of research. See for example Croft (2003) and Bower et al. (2011), which was published as part of the research

asks how whether tribes, which are the primary social unit in historical Arabic sources, can be considered speech communities and therefore whether they act as a linguistic unit in addition to their role as a social unit.

2.1 Modes of Subsistence

“If you see a cow, you are near a village, if you see a goat, you are near a camp, if you see a camel, you are lost.” — Baluch³ proverb cited in Barfield (1993, p. 95)

References to peoples known as Arabs began as early as the ninth century BCE in Biblical and Assyrian texts, then in Greek and Roman texts and later in Byzantine and Persian records (Retsö, *EALL*, “*Arab*”, 2002; Macdonald, 2009a; Hoyland, 2001, pp. 7–9).⁴ Arabs are often associated in these texts with desert dwelling nomads, and throughout these texts are generally referred to using similar names, suggesting that *‘arab* or something similar was a self-designation, especially given the relatively similar form of the term throughout the two millennia of attestation. This is not to say that all references to Arabs seem to refer to desert dwellers, and indeed Arabs are found in many different occupational roles (for a list of these, see Macdonald, 2009a, pp. 283–5), including as farmers and merchants. Indeed, the association of Arabs with a purely nomadic lifestyle may be a relatively modern

from the Dynamics of Hunter-Gatherer Language Change project: <https://webpace.yale.edu/huntergatherer/index.html>.

³ The Baluch are a largely single-elevation nomadic group in southern Iran that are largely sheep-goat nomads.

⁴ In the 9th century BCE, an attestation in an Assyrian cuneiform texts as *ar-b-a-a*, reconstructed by Retsö (2003, p. 126) as *šarba:y(a)*, in Herodotus in Greek around 440 BCE as *arábioi* (p. 244), *árabes* in Plutarch (p. 347), around the turn of the era, with the final term continuing in Roman sources.

misconception as a result of associating groups which practice a tribal social organization with groups which practice nomadism, though one certainly does not imply the other (Macdonald, 2009a, pp. 294-297)

Indeed, the methods of subsistence practiced by residents of the Arabian peninsula, who would have spoken a number of languages, including Arabic, were likely quite varied. Agriculture was clearly practiced in South-West Arabia, helping fuel the growth of the many empires (largely writing in Old South Arabian) which developed in this area, and on both coasts. King (1994) reviews a great deal of archaeological evidence that suggests that agriculture was relatively widely practiced in the interior of the Arabian peninsula, with complex systems of irrigation drawing water from groundwater at oases, allowing the growth of grains and dates. The agricultural evidence points to sedentary people living in tower houses (for protection from nomadic raids) growing a variety of crops in nearby fields. Many of these sites were located along trade routes, allowing residents to supplement their incomes as way stations. See Figure 2.1 for a somewhat schematic map of the extent of agriculture and oases in pre-Islamic Arabic.

Many of the residents of the Arabian peninsula, extending as far as the marginal steppe lands on the fringe of settled areas in the Levant and Mesopotamia, practiced varying levels of transhumanism, seasonal movement dictated largely by the needs of livestock. Donner (1989) recognizes three primary types of transhumanism in this area (my terms): altitude varying transhumanance, riparian transhumanance (the first two both small-stock transhumanism), and camel nomadism. Since altitude varying transhumanance, where the change of seasons at varying altitudes dictates the rhythms of migration, was relatively marginal among likely Arabic speaking groups, I will focus primarily on the latter two types. Note that groups may well change their style of subsistence over time, with a long

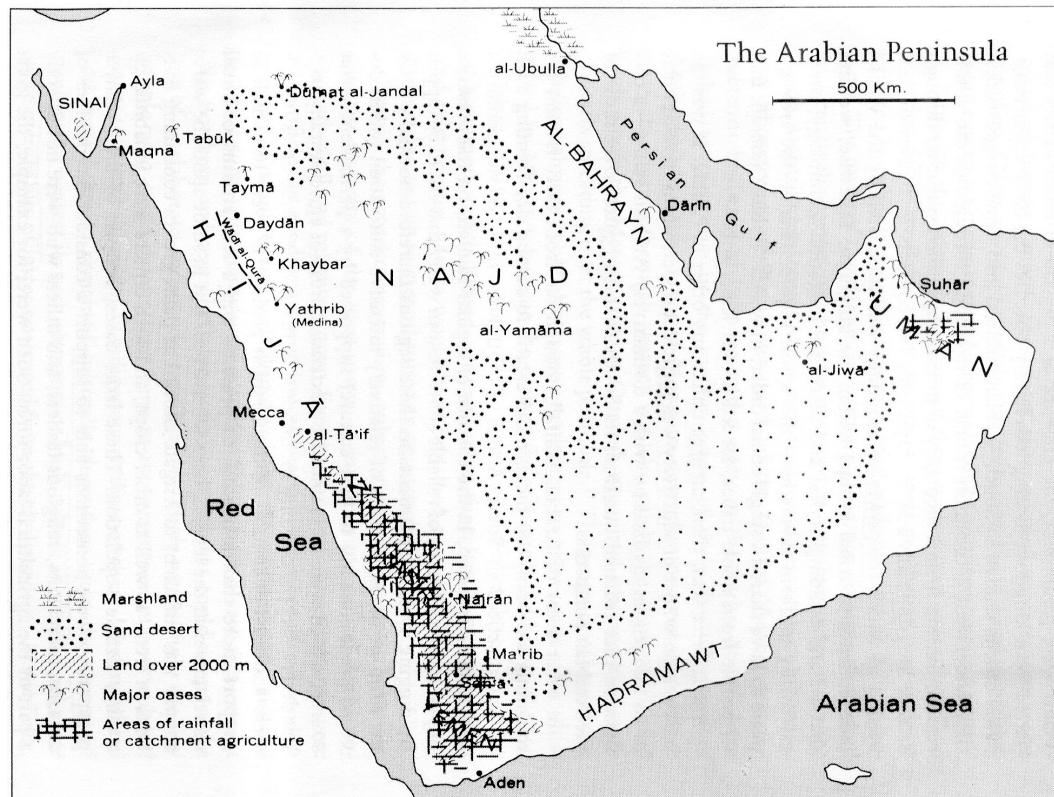


Figure 2.1: Schematic Map of Arabian Agricultural Areas, from Donner (1981, p. 13)

run of high rainfall enticing primarily camel nomads into raising mixed herds that include smaller stock animals, and periods of drought (and the raiding that can be result) pushing small-stock nomads who own some camels into moving further afield in search of pasture.

Riparian or river-valley transhumanance revolves primarily around raising small-stock animals such as goat and sheep, which require relatively frequent access to water and relatively soft foods. The relatively greater water needs of these small-stock animals (in comparison with camels) requires that these groups reside

relatively close to water sources throughout the year. In general, river-valley pastoralists move out into agriculturally marginal areas during the winter and spring, where even the limited rain in those areas permit the growth of food for the livestock, but in summer they necessarily move closer to agriculturally viable lands, normally near river basins (such as the Euphrates, Orontes and Jordan river basins) or in areas with relatively higher rainfall. This results in very strong ties between the pastoralists and sedentary peoples, since as Donner (p. 75) notes:

The riverain district is likely to be filled with villages and the pastoralists must keep their flocks well under control during the long summers to prevent them from ruining the villagers' crops. The pastoralist's fairly long stay among the villagers results in very intimate social ties between the two groups. Indeed, the two sometime become virtually one social group, part of which stays in the village year round and part of which takes the flocks into the steppe in the proper season.

Linguistically, these riparian nomads may be one of the major forces which move linguistic features along waterways. By moving between more settled groups, but having enough social ties with the groups they interact with, they could conceivably move innovations along a river corridor, such as the Euphrates or the Nile.

Camel nomadism, in contrast, exploits the ability of camels to move much further from water and to consume plants that other animals cannot, and therefore to take advantage of spring pastures much further from settled areas (and from competition with other forms of livestock). It is still, however, necessarily seasonal. Barfield (1993, p. 59) notes that milking camels in Arabian desert summer (with little forage) need watering once every four days, while in winter they can go four

to six weeks without drinking. This dictates a seasonal cycles where the hottest of the summer months must be spent near to water, typically around an oasis which would be occupied by settled agriculturalists. During these months, though the camels have sufficient water, they typically exhaust the nearby grazing and live from stored hump-fat. Once the heat has reduced from its peak, camel herders can roam widely, up to almost 2000 km in a season (Barfield, 1993, p. 66), though in the summer months movement is largely between wells (which groups may claim 'ownership' over, *ibid.*, p. 66). Groups can raise both camels and small stock, preferably goats which have lower water requirements than sheep, but in doing so they lose the advantage of the greater range of foraging made possible by camels.

The incredible long distance movements of camel-raising nomads means that they can potentially move between areas that speak very different language or dialects. Since these groups are not heavily integrated into the communities which they interact with, they are not as likely to participate in linguistic innovations originating in those communities. However, should they adopt mixed herds and dwell in an area, they may create a very different dialectal map than was previously found there, as detailed in Section 2.5 below. They may also be very inward facing, so two nomadic groups that share territories during part of the year are largely apart at least throughout the summer, and possibly for other seasons, and this may solidify linguistic divergence, as reported by B. Ingham, 1982 for the Āl Ḍafir and Muṭair Bedouin in north-east Arabia.

Pastoralists rarely raise their stock solely for subsistence. Indeed, the majority of their diet is grains (rice, bread, etc) and agricultural products such as dates, supplemented often times with dairy products (Barfield, 1993, pp. 69, 99). Camel pastoralists tend to consume milk fresh or relatively fresh from the camels, while mixed-herd pastoralists (who may also have camels) tend to produce more com-

plex dairy products such as butter or yogurt (dried or fresh) for consumption and sale. Most of the animals are raised for market, where they, and derivative products (wool, etc) are sold or exchanged for the agricultural staples that make up the majority of pastoralist diets. The advantage here is for small-stock pastoralists, since these smaller stock tend to reproduce and mature much more quickly, and whose closer ties to settled people allow them to participate frequently in settled markets. In many cases, small-stock pastoralism tends to be much more lucrative than agriculture in these marginal areas, and settlement can be tied to either great wealth (to the point where lands and herders can be bought or hired) or poverty, where wage-work becomes necessary when a flock is lost due to disease or famine (F. Barth, 1961; but see Barfield, 1993, pp. 118-126). Generally, successful pastoralists can diversify their holdings by acquiring settled lands, which are frequently worked by share-croppers, while the pastoralists continue to work largely as herding, especially since hiring labor often significantly reduces the profitability of herding (Barfield, 1993, pp. 118-126).

An unusual result of the mutual needs of nomadic and settled peoples is that nomadism makes far-flung settlement economically viable (*ibid.*, pp. 99-100). In many cases, nomads bring the market to the settled people, purchasing agricultural surplus that otherwise would need to be brought to market at greater expense, reducing the farmers' profits, while providing settled people with sources of protein, animal power (e.g. camels for caravan traders) and sometimes fertilizer (while eating crop stubble). This may be one of the reasons that desert oases were habitable at all, since bringing any surplus to market would likely be untenable without the mobile market provided by nomads. I explore the interactions between nomadic and settled groups further below in Section 2.3.

2.2 Group Organization

The social organization of Arabic speaking groups in pre-Islamic and, depending on the region, well into post-Islamic times (including the modern era), is often along the lines of tribal groupings. Tribes are, putatively at least, reflections of shared genetic ties. For example, Barfield (1993, p. 75), talking about models of lineage in camel nomads of the Arabian peninsula, focuses on the *faḥḍ* a ‘five generation lineage’ where all members are assumed to related by blood via an ancestor five generations prior. Marriage in those societies is typically endogamous to this group, though in practice it tends to be between first or second cousins.

However, even the notion of descent is fluid, and tribes tend to shift allegiances over time. This fluidity is reflected in the account by Bruce Ingham (1986, esp. Chapter 3), of the complex terminology for tribal groupings used by the Āl Ḍafir, a Bedouin camel nomad group in north-eastern Arabia. The name of the group itself has meanings related to interleaving and braiding, suggesting a composite nature that is acknowledged in the groups own oral histories — many of its members can trace their own descent to other lineage groups, though they owe their current allegiance to Āl Ḍafir. However, there are many different groups within the Āl Ḍafir, and these are not simply hierarchical. For example the *bayt* is a two or three generation grouping, while a *badīda* is “a much larger descent group often not necessarily traceable to a common ancestor but thought to be mutually related”. A *gabīla* corresponds approximately to the *faḥḍ* mentioned above, but is composed of multiple *badīda*. A *ḥilf* ‘federation’ can either describe confederate tribes, or confederations of tribes (who are not seen as related per se) which band together for specific purposes, for example mutual defense against raids by other groups (ibid., pp. 34-35).

In spite of the claims that these groups are united by mutual genetic de-

scent, both Barfield (1993) and Bruce Ingham (1986) emphasize the fact that blood relationships rarely actually dictates the self-described affiliations of tribal members. Instead, shifting alliances allow small groups to shift their allegiances to other groupings, and some units, such as the *ḥilf*, make little claim to genetic relation. The name of a group may differ based on the situation: Bruce Ingham (*ibid.*, p. 38) records the Āl Ḍafīr as saying of a subgroup Banū Ḥālid: *iḏā ġaw ḍufrān wiḍa rāḥaw ḥawālid* ‘if they come with us [as a result of migrations and camping], they are Ḍafīr and when they go away they’re Banū Ḥālid.’ Tribal groups can also absorb members from outside — Bruce Ingham (*ibid.*, p. 38) discusses a group referred to as Anṣār, originally Christian prisoners of the Ottomans in World War I who were later integrated into the tribal system through adoption and marriage, similar to the *mawālī* of the early Islamic era who were non-Arabs that were integrated into the tribal system.

It is therefore better to view tribal groupings not as reflecting genetic groups (in which membership is immutable) per se, but rather as corporate entities which form and operate in order to combine manpower and resources. The smallest component of this grouping would be the family (or perhaps the camping group), rather than the individual, but the definition of family is also somewhat mutable. An Arabic Bedouin saying goes, “me and my brother against my cousin, me and my cousin against strangers”, illustrating the complex recursive levels of identity that exist in any society. Solely as an analogy for readers in a modern industrial context, tribes in this sense they can be viewed in a way similar to that of labor unions, which form to advocate for the rights of their members, but in which membership is relatively optional and brings with it both benefits and responsibilities. Similarly, in an excellent turn of phrase, Hoyland (2009, p. 390) refers to them as the “political entities known as tribes”, a succinct way of indicating the primarily

political (and secondarily geneological) nature of these human groupings.

Not all of pre- or post-Islamic society was likely arranged in tribal groupings. Among camel nomads, tribal affiliation was extremely important for, among other things, maintaining the integrity of the camping group as well as for alliances of mutual defense (or offense) (for a complete list of reasons, see Donner, 1989, p. 80). In contrast, increasing settlement likely brought with it greater dilution of tribal ties in favor of other types of allegiances, such as allegiances with state powers, with groups aligned by trading interests rather than pastoral ones, or simply with the other groups in the city. This is reflected in the relatively smaller tents and tent groupings of small-stock nomads, which tend to be based around nuclear families rather than the extended families which share tents in camel nomadism (Barfield, 1993, pp. 74-76, 100-104)

That isn't to say that settled peoples lacked tribal affiliations altogether — as we have seen, in many cases with small-stock nomadism groups may simply be bifurcated, with part of the group tending the stock during the winter and spring migrations, and reintegrating during the summer months. Even when there is not this bifurcation, small-stock nomads necessarily interact frequently with settled peoples, and sections of these groups may sedentarize as a result of the economic pressures discussed above, again producing a split in a tribal group with spans the gap between largely settled and largely nomadic groups.

Similarly, even camel nomads may form either genetic or allegiance ties with variously more settled groups, especially given their need for both trade and for summer camping grounds. While the extreme mobility of camel nomads gives them the ability to raid settled peoples for sustenance, or even to exert power over them, the summer months dictate at least two to three complete months of immobility and it is better to have good relations with the settled peoples that occupy

the oases vital to the camel nomads' existence. Frequent raiding would reduce the viability of these groups, and remove a vital source of the staple foodstuffs consumed by the camel nomads.

Relations between camel nomads and more settled groups are also mutually beneficial. Settled people in an urban area who can draw on the military power of highly mobile camel mounted warriors can wield a significant source of power, even over those who can draw on the military power of established states. The hit-and-run tactics of camel nomads, who can strike and then fade back into the desert, are extremely effective in the areas directly adjacent to the deserts, though their ability to face organized standing armies in better watered areas was probably less impressive. For example, the Ghassanids and the Lakhmids were able to parlay their position on the edges of the settled areas of the Byzantine and Sassanian empires respectively into major dynasties by controlling the camel nomads of the deserts.

2.3 Group Interactions

In the surviving primary literature, the relationship between the nomadic Arabs and settled people (including other Arabs) is typically portrayed as antagonistic and in a great deal of the modern secondary literature. Nomads are portrayed as scoffing at the “servitude” of those dwelling under the rule of a state, as they lived in supposedly egalitarian tribal units. Sedentary peoples, on the other hands, considered Arabian nomads to be something akin to savage outlaws, unfamiliar with the benefits of civilization (apparently an ancient trope — see Shaw, 1982; Hoyland, 2001, pp. 96-96)

However, nomadic tribes also had the option of simply taking what they needed from settled people, who were quite vulnerable to hit-and-run attacks by

nomads. Settled people were easy targets, and nomads could always slip back into the deserts after their raids, making a military solution (other than the expense of maintaining a garrison or building fortifications) elusive. Even holing up behind walls was not terribly effective, as nomads could simply plunder or lay waste to the agricultural lands outside the fortifications until the owners agreed to pay tribute (Donner, 1981, p. 30).

Due to the relative ease of military domination by nomads over settled peoples, the balance of power tipped in favor of large nomadic tribal confederations, who would extract various forms of taxes (often closest in form to protection money) from the settled people in their dominions (*ibid.*). Indeed, Donner (*ibid.*) argues that large tribal confederations controlled by warrior aristocracies were the major powers in Arabia, with deep desert dwellers exacting tribute from semi-nomads, who in turn extracted tribute from settled peoples. However, it is important to recall that within a given tribe, there would have been branches that practiced a variety of subsistence modes, so that city dwellers could at times benefit from the military prowess of their nomadic brethren, as discussed above.

Also, it is important to keep in mind Hoyland's (2001) observations about the importance of actual subsistence — it seems unlikely that Arabia was a scene of constant struggle and warfare for dominance, as this would have interfered in the daily livelihood of everyone involved. Thus, on a day-to-day level (or more accurately, a year-to-year level) it seems more likely that interactions between the various inhabitants of Arabia were non-violent, relatively frequent, generally seasonal, with both short and long-distance trade occurring constantly, and the pattern of migration between particular watering holes and large markets bringing even nomadic peoples together on a regular basis. Warfare, raids, and pillaging would have been limited by necessity, since one can only raid so often before there

is little left to steal, or the victims of the raids move to a more secure location. The rule was probably an ordered symbiosis, as opposed to the romantic image of camel mounted warriors bearing down on poorly protected settlements and living off the booty.

Trade was a major source of income throughout the Arabian peninsula, and trade in goods from southern Arabia (particularly incense and other aromatics) and the Indian subcontinent dictated much of the geopolitics of the Arabian peninsula in antiquity. Empires such as those in south-west Arabia, and in the north such as those of the Nabateans and Palmyra flourished through the regulation of the international trade in aromatics, while Oman was also a valuable source of copper. The aromatics trade traditionally moved up the relatively accessible route of the Hijaz, with many settlements a short ride from one another, on the western coast of the Peninsula, though at various times it also moved into the Red Sea. Trade that moved east-west had to avoid as much as possible the Empty Quarter in the eastern part of the Peninsula, passing north of it, through Najd and Yamama to the central Gulf coast. Trade across the deserts would have required cooperation with the desert nomads, who might need to be paid for safe passage, while camel pastoralists were needed to provide the camels for these caravans.⁵ Major trading towns, such as Qaryat al-Faw, show how important trade was to the area, with much of the city arranged around the market and the caravanserai (al-Ansary, 1982). Local, low level trade would also have occurred between nomads and settled peoples.

Another source of power was the cities which contained religious sites, which could give rise to powerful dynasties (Donner, 1981, 34–37, ff.). These areas

⁵ Here the practice of raiding caravans might be more lucrative than symbiotic subsistence, but it might be hard to sell the goods you had just stolen back to their original owners.

were sacrosanct, tradition forbade fighting these places spiritual centers. They thus emerged as safe places for negotiations between warring groups, and eventually as important market centers. The function of these enclaves was of such value that they were more important in this role than as captured holdings, and they therefore enjoyed a great deal of autonomy and political power. Obviously the best example of this type of settlement is Mecca, with the Ka'ba, the cultic shrine that is now central to the Islamic hajj pilgrimage, but which had religious significance even in pre-Islamic times. Mecca's role as a cultic center allowed it to become a prosperous trading city and gathering place for tribes coming to worship.

Major settled empires sought ways to mitigate the danger posed by nomadic groups. One of the most common tactics was to form an alliance with a semi-sedentary or nomadic group which, with the resources provided by the settled empire, could then dominate the nearby nomadic groups (using military might or bribes from the deep purses of their sponsors) and curtail their raids. These groups would also collect taxes, and provide military auxiliaries for their sponsoring empire. The empires of Byzantium, Himyar and the Sassanians all employed this tactic (Donner, 1981, 37–49, ff.). It is not clear exactly what subsistence strategies these groups followed prior to taking on this role, but they generally became settled to various degrees after developing this alliance. The Ghassanids, working for the Byzantines, built monumental churches and left significant archaeological remains (with many inscriptions in Greek). The Lakhmid dynasty was similarly centered in the religiously important town of al-Ḥira, one of the major pre-Islamic Arab cities.

The types of group interactions that existed in pre-Islamic Arabia would have had linguistic consequences since they would dictate the social network links between groups, as well as the allegiances held by the peoples involved.

At first glance, it would seem logical that the unequal relations between nomadic and settled people would place the nomadic camel warriors in a dominant position over the settled people, which we would expect to make the speech of the nomadic group more prestigious than that of the settled group. However, for linguistic features to diffuse, there would need to be enough network ties between the two groups — after all, there is not much time to talk during a midnight raid. A truly nomadic group would have very little success imposing its language on nearby settled populations, simply because there would not be enough contact. Moreover, social stigma which settled people held towards nomadic groups could also prevent the movement of linguistic features, since the settled people would not feel that they were linked to the nomadic groups by a strong bond of allegiance.

It seems more likely that the groups which operated between the major settled empires and the nomads would be more important as centers of linguistic prestige. These groups were largely settled, but had important links of alliance with nomadic groups. We know that control of these buffer kingdoms was desirable and contested — we know of two groups which seized control of the Levantine buffer kingdom in succession, the Salihids and the Ghassanids — and their relatively great material wealth, backed up with the military and financial support of their sponsor, would have made them major powers in the region. At the same time, of course, even alliance can be grudging, and clearly these groups owed at least some of their power to overt subjugation of some other groups. Among close allies who saw themselves as part of the same federation, there might have been direct diffusion of linguistic features. Among groups that felt some need to differentiate themselves from the ruling group (or were differentiated in fact by a subordinate status) there might still be some accommodation to their linguistic behavior, but possibly through the use of calquing or of certain shibboleths while

allowing some features to diffuse.

Trade and pilgrimage would also have diffused features in a hierarchical manner, first between settled areas, and then possibly into the surrounding regions, with changes diffusing out of relatively more important centers. Pilgrims would be similar to the soccer fans and tourists which bring London features into Norwich. Trade centers would probably have been linguistic prestige centers also, with the many groups necessarily visiting those centers and interacting with their wealthy and powerful residents.

It is unclear whether those groups who provided the transport would themselves have been linguistically effected by the language of their clients; certainly they probably needed some amount of linguistic flexibility to interact with people at various stops along the caravan trade.⁶ Their clients were probably the main diffusers of linguistic features from one settlement into another, as they would have relatively important ties and group affiliations in both their origin and their destination, and therefore we would expect similar linguistic behavior in major trade centers.

2.4 The Tribe as a Linguistic Unit

It has been frequently assumed that, in addition to a political unit, tribes form a linguistic unit, both in the early Islamic grammatical literature and in much modern linguistic literature as well. This is especially evident in the literature based on early grammarians' descriptions of languages which attribute certain linguistic traits to particular tribal groupings (Fück, 1950; Al-Jundī, 1983; Rabin,

⁶ One wonders whether anything like the Mediterranean *lingua franca* emerged in this environment, with trade between groups speaking Greek, Persian, and various Semitic languages.

1951), which leads to a mass of contradictions in the data (al-Sāmarrā'ī, 1994), with the same feature being attributed to multiple distinct groups, or with tribal names apparently acting as proxies for geographical areas (hence Rabin's more general attempt to distinguish eastern and western dialect areas).

The problem is one of size of analysis, and centers on the fact that very few of our sources give us any indication of the size or cohesiveness of the groups that they refer to as "tribes." Many of the tribes that we hear of in chronicles of Arab conquests are clearly massive confederations — Tamīm and Ṭayyi' both clearly covered vast geographical spaces (Ṭayyi' were so widespread, their name was used generally to refer to nomads in Syriac), while the Prophet Muhammad's tribe of Qurayš seems to have been relatively small (and concentrated geographically). The Ghassanids and Lakhmids interacted with many other named tribal units, but the tribes who were both subordinate to those groups and living within that geographical area probably shared linguistic features in a way that probably corresponded to geographical location more than tribal affiliation.

We cannot, therefore, take for granted the idea that the tribal unit in the early Islamic historical sources (or even later sources) was equivalent to a speech community. That is to say that a tribe was not necessarily a linguistic unit. An entity referred to as a tribe in those sources could have been either a very large or very small group of people, and only smaller groups were likely to form a cohesive speech community due to the difficulty of maintaining network connections. Larger groups were more likely to contain members pursuing a variety of different subsistence patterns, and therefore would be unlikely to have regular contact between members.

For example, we know that for many tribal confederations, there were both sedentary (or largely sedentary) members, and members who spent large quanti-

ties of time isolated in pasture lands. The sedentary sections of the tribe would likely participate in the wider diffusion of features between urban centers, while the more nomadic elements would likely not participate in the same innovations (see also below, Section 2.5). Moreover, looking at modern maps of Middle Eastern migrations (TAVO, *Nomadismus und anderer Formen*), we can see that a number of modern tribes which share the same name are separated by great distances, and do not share territory at any point in their migrations.

Moreover, as we have seen, the definition of a tribal grouping tends to be a matter of political convenience rather than a socio-cultural determinant, and these definitions tend to be fluid over time as alliances shift and change — there are numerous examples during the early Islamic conquests of tribes split politically “sometimes along lineage lines⁷, sometimes along religious lines, and sometimes along lines that cannot be ascertained (Donner, 1981, p. 182).” Obviously community allegiances can be multiple and overlapping, but in this case the “tribe” does not appear to be a dominant identity.

We would expect, therefore, that desert dwelling groups, who meet and interact at wells throughout their migrations, would become linguistically similar (if they were not already), and that this would cut across tribal boundaries. The “[winter pastures are] occupied by an assortment of different tribes, and there is considerable visiting and feasting between them (Barfield, 1993, p. 67)”, while it is only the three months or so spent at the summer camp when an entire lineage is united. Small-stock nomads, on the other hand, probably interact more with settled groups, but still interact with each other more than they do with either settled groups or camel nomads. Since tribal boundaries can rarely be said to coincide

⁷ That is to say, a level of analysis below the level of named tribes normally available to us in the sources.

exactly with settlement pattern, the tribe is therefore unlikely to define the extent of linguistic network connections, even if it could be said to reflect allegiance. Again, this is not to say that the tribe cannot be a speech community, but rather it cannot be assumed that all units referred to as “tribes” are necessarily speech communities.

Marriage practices also seem to ensure the dominance of subsistence pattern as a marker of identity over lineage. Lineage group endogamy is not actually that strong. While it is true that Bedouin groups tend to marry within the lineage, loosely defined, actual first cousin marriages — the culturally ‘ideal’ marriage arrangement — are quite rare, with second cousin and more distance marriages are indeed the majority. Marriage among more distant kin gives alliance advantages, almost like a true out-group marriage (Barfield, 1993, pp. 76-78). On the other hand, many Bedouin groups have strong taboos against marrying members of groups which follow other subsistence patterns. The Negev Bedouin, for example, have taboos against nomadic herders marrying ‘down’ to settled people that may be from the same tribal lineage (Henkin, 2010, pp. 40-41). The primary exceptions to this is for men to marry women from settled groups, though the women must move and live with the man’s family in a patrilocal system, so that the children are raised in a nomadic lifestyle. Linguistically, this would be reflected in divergence between the settled and nomadic groups, even if they were putatively from the same tribe.

Why then do linguistic sources, both ancient and modern, treat the tribe as if it *were* a linguistic unit? There seem to be a number of possible answers to this question. The first is that it seems an easy assumption to posit a relationship between the unit of social groupings — the tribe — and manifestations of social groups (i.e. language). Without quality ethnographic research, this assumption

is an extremely tempting one to make and seems *a priori* valid, while in light of ethnographic data it is much less defensible.

There is also an illusion of tribal cohesion that results from the fact that tribes did act as a political unit. Thus, tribes moved into the newly established cities in Iraq, for example, and tended to move as a unit. The city of al-Kūfa carefully preserved the tribal divisions of its settlers, with each tribe given its own relatively independent quarter of the town Donner (1981, pp. 226–237). When new members of some tribes arrived and overwhelmed the area allotted to their tribe, the entire tribe moved outwards from the center of the city rather than being split, suggesting that tribal ties were still extremely strong at this stage.

However, it turns out that these ties were not so much cultural or linguistic, but rather financial/legal (and possibly political). As Kennedy (1998a, p. 62) notes, “because payment was made to the tribesmen through their chiefs (*ašrāf*) members of tribes tended to camp and settle in the same general areas,” that is the political unit was also the financial unit. Over time, this unit may have become a more homogeneous linguistic one, but since this was in the context of a larger city, it probably tended to assimilate to the language of the city generally rather than the tribe specifically. Moreover, in the case of Kufa, under Mu‘āwiya (ruling from 661–680) the city was reorganized into quarters, suggesting that a purely tribal division might no longer have been as relevant at this time, a rather short time after the founding of the city in 637–8.

Another potential reason for this confusion is the nature of linguistic description in the early Islamic tradition. Arabic linguists living in large cities in the eighth and ninth centuries would make use of consultants (who provided them with largely archaic, literary texts, primarily poetry). Though it is a common statement that such-and-such a linguist spent time in the desert learning the true

Bedouin speech, this appears to be little more than a trope (and possibly one which was part of a larger ideologizing process to valorize desert nomadism). Moreover, even if they did work with an informant who happened to belong to tribe X, attributing the linguistic behavior of that informant to the remainder of the tribe would be a common mistake, and many of the contradictions in the data discussed by al-Sāmarrā'ī (1994) may well arise from different grammarians repeating this error with different informants who might belong to different tribes but which were part of the same linguistic grouping (especially since these informants were close at hand to the cities of southern Iraq.)

Moreover, the ephemeral and changing nature of tribal alliances might well mean that a linguist, ancient or modern, may well mistake the place of an informant within the tribal structure. Barfield (1993, pp. 74-76) describes the difficulties others have in understanding what actual group an individual, or even a subgroup belongs to. Those who have been attacked by what they perceive to be a member of a named tribe (Rwala, Shammar, etc) might complain to members of those federations, only to be told that the attackers are part of an only loosely affiliated group (potentially creating misguided retaliation). Even ethnographers, trying to describe what are actually quite mutable groupings, might create an illusion of certainty that does not actually reflect the reality of the alliances and links between nomadic groups.

This study, therefore, rejects the notion that tribes formed a linguistic unit, and in reconstructing dialect groupings, any attempts to tie these groupings to social units will focus on geographic and subsistence pattern groupings rather than to named tribal groupings.

2.5 The Bedouin Paradox

A frequent trope in the literature on Arabic is the notion that nomadic groupings are somehow repositories of more ancient values and linguistic forms. The very notion of using Bedouin informants in the early Islamic linguistic literature reflects this trope, as do the various adjectives associated with Bedouin speech — ‘pure’, ‘uncorrupted’, ‘eloquent.’ Even the quite ethnographically aware article by Donner (1989) falls into the trap of viewing Bedouin as more conservative than settled groups, with respect to two features — the maintenance of the tribe as a social grouping (an obviously functional result of subsistence pattern), and linguistic conservatism.

This apparent linguistic conservatism, however, is primarily the result of the differential contact between nomadic groups and settled groups, versus the contact between different types of settled groups. Linguistic changes tend to move along the urban hierarchy, typically with high prestige urban variants of linguistic variables spreading to increasingly sparsely populated areas, though movement can occur in the opposite direction. Nomads are even less likely than even rural settled people to participate in this urban hierarchy, and this distinction can be sufficiently pronounced to appear as a qualitative (rather than simply quantitative distinction).⁸ Thus, if an innovation has diffused through the urban hierarchy, and not yet appeared in Bedouin dialects, this creates an illusion of conservatism in Bedouin dialect vis-a-vis the urban centers.⁹ Palva (*EALL: “Dialects: Classification”*).

⁸ The general trend is best expressed simply as a the differentiation of cities from less populated areas, creating linguistic islands. In some areas, this has led to both an urban-rural distinction, in addition to the settled-Bedouin linguistic distinction. See examples of this in Palva (*EALL: “Dialects: Classification”*).

⁹ There is also the possibility that the tight-knit communities in which nomads find themselves, with a small number of strong network connections, may well be more resistant to the spread of linguistic innovations versus the many weak network connections that many city dwellers likely

tion”, p. 606) is sensitive to the illusory nature of this conservatism, and helpfully lists a number of innovative linguistic forms found in Bedouin dialects, such as the reanalysis of the /n-/ from the passive N-Stem verbs (*infa‘ala*) to verb forms with the reflexive /t-/ prefix. C. Holes (2006) similarly considers the /g/ realization of the Q variable, widespread in Bedouin dialects, to be a relatively recent innovation.

At the same time that these Bedouin dialects are resisting linguistic changes arising in nearby urban centers, they are also highly mobile and able to sweep into areas to which they have not previously been native. The low population density of steppe-lands in which nomads dwell make it relatively easy for a new Bedouin group, possibly speaking one dialect, to replace the previously spoken dialect in the area (Buchheit, 1988; Ostler, 2005). Bedouin settlements in the Middle East have been reported as having population densities ranging from 2/km² to 15/km² in the modern era (George, 1973), which would be concentrated in small camping groups. If one camping group moved out of an area, and another, speaking a different dialect, moved in, from a perspective of ‘dialectology’ it could represent a significant change in the of an area, when really all that is changing is a very small-scale movement of people. In contrast, even small villages are difficult (and unlikely) to move in such a whole-sale manner, and cities are certainly much more stable over time in terms of their (apparent) linguistic behavior.

These two aspects of nomadic speakers — their resistance to the spread of innovation and their extreme mobility — conspire to create what I will call the Bedouin paradox:

THE BEDOUIN PARADOX Nomadic speakers generally do not participate in the

have (L. Milroy, 1980). Or they may have more weak ties with more different social groups, and be more likely to adopt new innovations. However, it is not clear how this actually unfolds over time and this is largely speculative.

spread of innovations among settled groups, and therefore they appear to retain archaic linguistic features in comparison with their settled neighbors. However, their extreme mobility and the easy of replacing indigenous nomadic groups means that these 'archaic' speakers may be newcomers to an area in comparison with settled groups.

This suggests that settled groups often reflect more ancient, in-situ dialects, while Bedouin nomads often bring with them dialects that are new to the area. The steppe is an excellent example of a spread zone (in a somewhat more micro-level of analysis than was intended by the term in Nichols, 1992), while in the short term cities, subject more heavily to the principle of first effective settlement, act as residual zones. The development and diffusion of innovations between cities and then down into the countryside create the illusion that the city dialects are somehow less archaic than Bedouin dialects, but there is probably greater long-term continuity in the dialects of cities versus the steppe.

This implies that Bedouin dialects are subject to rapid replacement over short periods of time. The modern Bedouin dialects of Syria or Iraq are by no means the same dialects as those that were present a season, a hundred years or a thousand years ago, even if those dialects appear to be 'archaic.' It is possible that Bedouin dialects previously present in those areas were more similar to the sedentary dialects, but were replaced recently by apparently more different dialects. It is also possible that due to economic conditions, semi-sedentary groups become less sedentary, and then migrate, and bring with them the sedentary linguistic features of their old home into an entirely new area.

At the same time, an accretion of Bedouin features may accumulate in the dialects of those settled people closest to the steppe, as successive waves of nomads

move into an area and influence the language of those they interact most closely with (possibly also a result of a transition from a nomadic to a more settled life.) Dialects just past the steppe may potentially be an interesting type of residual zone, where the waves of Bedouin leave behind a residue, like the seafoam (and flotsam) left behind by the edges of waves along a shore.¹⁰ However, the low population density of marginal settled areas might also mean that any relics would be quickly swept away by the next wave of nomadic influence.

2.6 Conclusion

This chapter has explored the relationships between groups following different modes of subsistence in the pre-Islamic Arabian Peninsula. It has shown that there are complex ties of mutual interdependence between nomads and settled people, and that stereotypes of constant warfare between these groups have been exaggerated. The balance of power in the pre-Islamic Arabian Peninsula would have been complex, shared between militarily powerful (but numerically weak) but resource dependent nomadic groups, resource rich but difficult to defend cities and oases, and cultic centers which were able to provide safe spaces for groups to interact and trade. This balance of power was often tipped by the major imperial powers in the region whose lands were adjacent to the steppe, and who used their resources to empower some groups over others, creating principalities which straddled the steppe and richer lands.

In terms of the linguistic implications of pre-Islamic social organization, this chapter has shown that the notion of a tribe is too vague and multivalent to

¹⁰ Donner (1981, pp. 95-6) uses a similar sea analogy to demonstrate how a villager living on either side of a desert would be no more able to cross that desert than to cross an ocean, unless they had the expertise of a sailor or desert nomad.

be taken as synonymous with a speech community. It has shown that there are a number of reasons for the divergent linguistic behavior of settled and nomadic peoples, particularly the diffusion of features along a hierarchy of population density, and the extreme mobility of nomads. This results in an apparent gulf in the linguistic behavior of settled and nomadic people that is as much the result of population movements as the diffusion of linguistic features. It has argued that in spite of nomadic military superiority in some areas, these nomads were not likely to be the determiners of local linguistic behavior. Instead, the semi-settled border principalities were more likely to set the linguistic tone for an area. It has also shown that some nomads, such as small stock nomads who are in frequent contact with settled people may play an important role in the diffusion of linguistic features along marginal lands at the edge of the steppe.

Chapter 3

History: Greater Arabian Peninsula

The goal of this chapter is to understand the history of Arabic speaking peoples within what I term the GREATER ARABIAN PENINSULA. This term refers to the essentially contiguous lands stretching from modern Oman and Yemen in the south to the Levant in the northwest and modern Iraq in the northeast. The boundaries of this region are the bottleneck of the Sinai peninsula in the west and the segments of the Alpide belt of mountains that goes from south-western Anatolia (the Taurus range) into northern Iraq and Iran (the Zagros range).¹ The rest of the Peninsula is bounded by water — the Red Sea in the West, the Indian Ocean to the south, and the Persian Gulf to the east. A map of the land boundaries is shown in Figure 3.1.

This area obviously has its own internal divisions, including the Empty Quarter desert in Saudi Arabia, the Euphrates and Tigris Rivers, and mountain ranges in Lebanon and Yemen. However, it does form a geographically contiguous area, and within this geographical area, Arabic is still largely confined to the

¹ Bosworth (1983, pp. 603-604) notes that north-eastern Oman is actually divided from the rest of the Peninsula by the same range of mountains that form the Zagros, and indeed that part of Oman has had a history separate from the rest of the Peninsula, with a tendency to be controlled by powers across the Gulf rather than from within the Peninsula.



Figure 3.1: Map of the Greater Arabian Peninsula

Greater Arabian Peninsula region — even within the borders of modern Iran, Arabic spread only to the base of the Zagros mountains, and in the north Arabic did not move past the Taurus range. In the immediate post-Islamic era, the Greater Arabian Peninsula was the center of Arabo-Islamic culture and power, and remains so today, though Egypt has gained greater prominence over time.

There are two primary goals for this historical overview. The first is to develop hypotheses about the pre-Islamic linguistic situation in the Arabian peninsula, especially the distribution of Arabic and other languages, and how speakers of those languages interacted. The second goal is to understand how the Islamic conquests altered this linguistic balance and ultimately lead to the Arabization of most of the Greater Peninsula. Understanding these historical changes is essential to understanding how the Arabic language developed, and is key to reconstructing pre-Islamic speech communities in Chapter 6.

3.1 Approach

Since this chapter is not the focus of the dissertation, and because it is not meant to bring anything new to the history of this region per se (except perhaps some new perspectives), it is a largely synthetic overview of the historical research that has already been done on this subject. Little attempt has been made to make use of the primary sources, except where there is some dispute over them by historians, and the reliance has been primarily on secondary sources. The goal is not to simply recite the various accounts of the pre-Islamic world and the Islamic conquests, which have already been covered in great detail (most recently and comprehensively in Kennedy, 2007); instead, the goal is to look at this history from a specifically linguistic perspective.

Though the Middle East and North Africa are now largely Arabic speaking, this is the result of a long-term, and in many cases still ongoing, process of Arabization.² It would be a massive undertaking to trace the full history of the spread of Arabic in each of these countries. This is simply impossible in the context of this dissertation, but it is not actually necessary for our purposes.

Instead, it is possible to focus on the establishment of the first VIABLE LINGUISTIC COLONIES in each region. A viable linguistic colony is a settlement occupied by speakers of a language that is sufficiently important and resilient to a) maintain the language, and often to cause nearby speakers of other languages to shift to using the new language as a primary spoken language of daily use. The term "settlement" is intentionally vague - a viable linguistic colony may be a physical, fixed location, such as a city or a quarter of a city, or it could be embodied in a nomadic group that changes location over time. Its defining property is that it be

² By Arabization, I am referring to the uptake of Arabic as the primary spoken language of daily interaction, usually with the loss of previously dominant languages.

sufficiently robust to maintain its language over a long period of time. By the principle of first effective settlement, we can expect these viable linguistic colonies to generally determine the language of a region, with shift to another language only happening due to fairly extreme circumstances.

Thus, instead of attempting to cover the whole of Middle Eastern history, we can restrict ourselves chronologically to the point at which viable linguistic colonies of Arabic speakers have been established, typically within the first two centuries after conquest. Once they are in place, we can assume that the shift to Arabic was a long, slow processes that likely included the establishment of other linguistic colonies. However, since these other linguistic colonies were populated from the same area, and because linguistic traits spread generally areally, the linguistic signal of the different phases of Arabization (following the first major phase) is likely too weak to be detectable.³

Following the establishment of the first linguistic colonies, we can assume that a process of Arabization occurred which would have slowly led to the shift from their native language to Arabic. Speakers nearby and in similarly sized settlements would probably shift first, while those in far-flung or geographically inaccessible areas (mountains, etc) would have been slower to shift. These shifts would have taken generations, and there would have been a variety of different shift scenarios: those non-Arab speakers living in Arab dominated areas would probably acquire Arabic in late childhood or early adolescence from their Arabic speaking peers, leading to early, high proficiency bilingualism. Further away from these

³ This argument is not based on the assumption that these areas were somehow “destined” to be Arabized. Instead, it is based on the historical fact of Arabization in specific regions, and hopes to establish how this came about, based on a combination of linguistic principles, linguistic evidence, and extra-linguistic evidence such as social history. The continued geographic spread of Arabic in countries that are not entirely Arabized is certainly not assured, and with movements towards schooling in some minority languages it may actually be halted or reversed.

centers, Arabic would probably have been acquired by older adolescents or adults for instrumental purposes such as trade. As Arabic-speaking groups expanded, there would be more opportunity for acquisition of Arabic in early adolescence from direct peer interaction.

Once we have some sense of where Arabic was spoken, and how different groups expanded out of the Greater Peninsula, we can then link the varieties spoken today to earlier dialects by concentrating on what dialects formed the bulk of the speakers in the linguistic colonies. When we perform the linguistic reconstruction, we can then link the reconstructed groups to the original colonizing groups, and have some notion of the dialect geography of the pre-Islamic Arabic-speaking world, as we do in Chapter 6.

I also want to draw a distinction between political history, the history of the ruling political class, and cultural history, the history of the human culture of an area (mentioned briefly, in the context of archeology, in Zeyadeh, 1994). The two are often discontinuous — a change in rulers very rarely results immediately in a change of culture. Language is also quite slow to change, and normally requires huge changes in population demographics.

However, changes in political rulership may lead to other cultural changes, such as the language of prestige and government, religion, the orientation of trade, and the invisible state borders or frontiers that separate different groups. In the case of the Islamic conquests, the language of prestige (learning, administration, religion) shifted from Greek and Aramaic to Arabic, with the dominant religion shifting eventually to Islam, the orientation of trade shifting from the oceans (and hence to Anatolia) into the deserts (El'Ad, 1982), and the previously permeable boundary between Anatolia and the Levant became a frontier and a war zone. These changes undoubtedly resulted in subtler changes, such as the allegiances of peo-

ples, notions of membership in certain groups, and evaluations of language and language use, though we would be hard pressed to find firm evidence for this. In Iraq, the religion of the Magians, though it had been on the decline pre-Islamically, was dealt the coup-de-grace by the Islamic conquests (see Section 3.6 below), though many of the cultural practices that Magians had engaged in continued for a long time (see Morony, 1984, pp. 201-2).

Kennedy (2007, p. 376) succinctly sums up the relationship between the change in political leadership and human cultural behavior:

The Islamization and Arabization that followed conquest over the next two or three centuries would not have occurred if political conquest had not already succeeded, but they were not a direct and inevitable consequence of that conquest. Instead, it was a gradual, almost entirely peaceful result of the fact that more and more people wanted to identify with and participate in the dominant culture of their time.

3.2 General Trends

This section will survey some general tendencies which are true regardless of region.

3.2.1 Subsistence Patterns and Barriers

As described in Chapter 2, Arab-speaking populations inside and outside the Arabian peninsula would have pursued a variety of subsistence styles, and the places in which they settled would have reflected that. There was probably significant Arabic-speaking settlement in the Hijaz for a long time preceding the Islamic conquests, and camel nomads would have exploited most of the desert

areas in the peninsula, though we cannot of course be entirely sure they were Arabic speaking. These desert areas are essentially continuous, stretching north out of the Arabian peninsula proper into the area delimited in the east by the Euphrates, and in the west by the increasingly fertile lands of the northwest Levant. The extent of the desert fringe can be seen fairly clearly in Figure 3.2, though it is important to remember that rivers allowed for irrigated agriculture in spite of low rainfall.

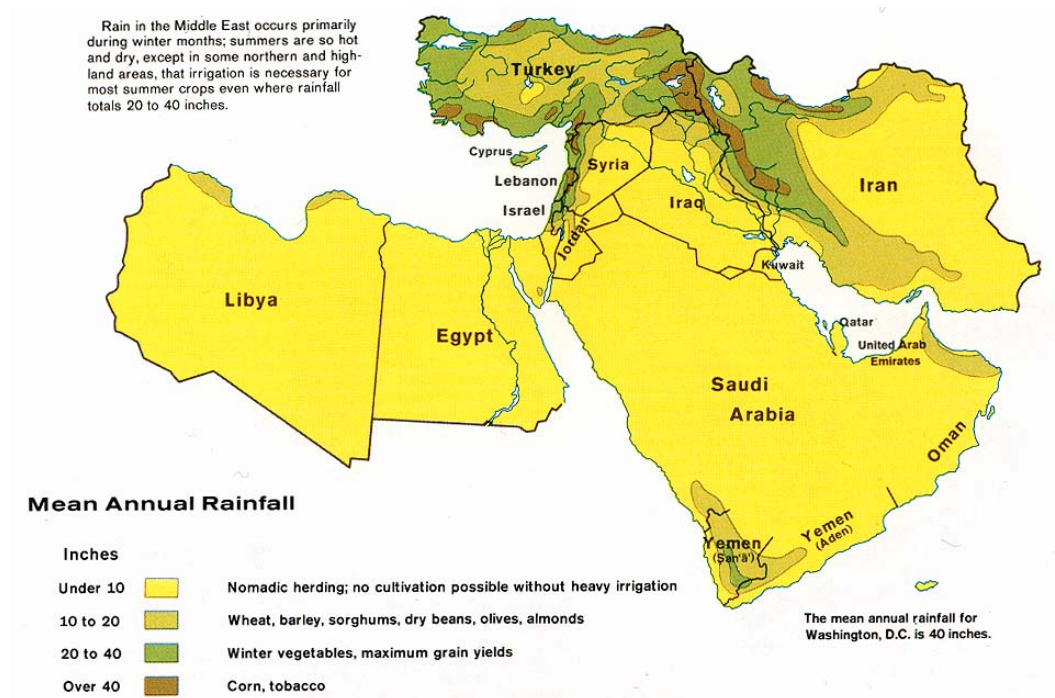


Figure 3.2: Map of Rainfall in the Middle East, from CIA (1973), retrieved from http://lib.utexas.edu/maps/middle_east_and_asia/middle_east_rainfall_1973.jpg

We can, in a sense, view the line that separates low water areas from better watered areas as a kind of barrier to the movement of nomadic pastoralists.⁴ This

⁴ Normally this line is that separates areas with more or less than 200mm/year of rainfall, though

is not to say that somehow their stock would not do well in better watered area, but rather they would come into direct conflict with those groups already resident in the more fertile areas who would already be exploiting pasture lands for growing and grazing. Rivers could at times act as similar barriers, since these densely farmed areas could be resistant to incoming nomads, though there might also be symbiotic relationships.

The rain line certainly acts as a linguistic barrier, since population density tends to correlate with rainfall or water sources. The difficulty of colonizing an area linguistically increases with the native population density, and when settling even in a less densely population area, it is likely that the dominant language of the region would eventually be adopted by an immigrant group. Nomadic pastoralists would be more resistant to assimilation due to having fewer ties with their neighbors, while semi-settled groups would be more likely to assimilate linguistically, and groups moving directly into a dense urban space would be very likely to assimilate within a few generations.

Since almost all of these high-rain areas were thoroughly settled and controlled by non-Arabic speakers prior to their settlement by Arabs, we can view the 200mm rain line as marking, with some caveats, the extent of the distribution of Arabic prior to the Islamic conquests. There was definitely some Arabic speaking settlement past the 200mm line, and political events changed this also (as with the penetration of Arabs into modern Yemen, or the establishment of the Arabic-speaking phylarchs in the Hawran), but this works as a very general rule. It seems to be rather accurate, since most of the early Arabic, pre-Islamic inscriptions in the Levant lie quite close the 200mm line, usually just outside it.

some agricultural can be practiced with sufficient catchment systems, as described for Yemen by Dresch (1989).

There are reports of Arabs living in more fertile regions, such as the Beqaa valley in Lebanon and along the Nile in Egypt, but it is not clear if this is really a linguistic designation or an ethnic one. Macdonald (2009a) has done an excellent job of collecting reports of Arabs in historical texts, and argues that the term ‘Arab’ must have referenced a ‘ill-defined complex of linguistic and cultural characteristics, (p. 319)’ though stereotypes and ‘topoi’ developed around the term.⁵ However, we know that ethnic categories tend to be longer lived than linguistic behavior, so a description of someone as Arab (or even a self-description) does not necessarily entail anything with regards to their language use, any more than someone claiming Italian roots in the US can be expected to speak Italian. At times, there were also legal incentives to maintain the ascription of Arab — for example, in Egypt, where Macdonald finds the most instances of self-identification as ‘Arab’ in inscriptions, Arabs payed higher taxes than any other group during the Julio-Claudian era (ca. 44 BCE — 68 CE) (*ibid.*, fn. 12).⁶

Given what we know about language behavior, we would expect that even if these ‘Arab’ groups did speak an ancestor of Arabic, once they became settled in areas that were predominantly Aramaic or Coptic speaking, they would be likely to assimilate linguistically to their dominant neighbors. The result of this tendency would to create an apparent congruence between nomadism (whether camel no-

⁵ Most of the reports of Arabs, excepting those in Ptolemaic Egypt, do indeed correspond with both the Greater Arabian Peninsula and the 200mm rain line, though sometimes Arabs are reported further inland, with Pliny describing Arabia as starting near Cilicia in the northwest corner of greater Syria (*ibid.*, fn. 20). Similarly, there is one ambiguous description of Arab garb being worn in Central Iran, but this account is very unclear (*ibid.*, fn. 28).

⁶ In Ptolemaic Egypt, Arabs appear to have been a largely settled group, practicing agriculture, and were seen as part of the native population by the Greek and Roman rulers. Given our linguistic expectations, it seems likely that they would have assimilated to the local Coptic language (or even Greek, the language of the inscriptions) over time, and that ‘Arab’ would have operated primarily as an ethnonym or administrative designation.

madism or small-stock semi-nomadic lifestyles) and the use of Arabic — only nomadic or fringe groups would be likely to maintain Arabic as a language of daily life, while others would assimilate linguistically even if they might retain an ethnic identity. This is a different argument from those criticized by Macdonald (2009a,b) that assume that all Arabs were necessarily camel nomads.

3.2.2 Linguistic Separation

There was a clear concern on the part of the Muslim conquerors about becoming assimilated by their newly conquered subjects. The literary tradition preserves various examples of provisions aimed at maintaining a separation between the Muslims and their new subjects. “Do not learn the barbarous language (raṭāna) of the foreigners (al-ā‘ājim),’ ‘Umar b. al-Khaṭṭāb is said to have ordered (Tannous, 2010, p. 512)” His successor, ‘Umar II, is said to have forbidden the sale of wine in the new cities, and required that any churches in these linguistic and cultural beachheads be destroyed (*ibid.*, p. 522). Even basic social functions such as greeting non-Muslims with a handshake or eating with non-Muslims were circumscribed (*ibid.*, p. 541). It is not clear whether the majority of people actually followed these rulings by religious and political authorities, but they indicate the fear of, and attempts to minimize, linguistic and cultural assimilation: “the dress rules (etc.) for non-Muslims can also (or only?) be understood as protecting the Muslim conquerors who began to settle in the conquered territories, but who were still a small minority in an alien environment (Noth, 2004).”⁷

In the pact that was supposedly concluded between the first Umayyad

⁷ Indeed, intermarriage, especially of non-Arab women to Arab men was quite common and there certainly were many commercial dealings between Arab-Muslims and people from other groups — given the demographics, it would probably be impossible not to.

caliph ʿUmar b. al-Ḥaṭṭāb and the conquered people of Jerusalem, the conquered people agree that:

We shall not seek to resemble the Muslims by imitating any of their garments, the qalansuwa, the turban, footwear, or the parting of the hair. We shall not speak as they do, nor shall we adopt their kunyas (M. R. Cohen, 1999, p. 107) ⁸

This restrictions were attempts to “erect [...] boundaries to discriminate between the conquered and the conquerors – and thereby reinforce the separate identity of the latter. (*ibid.*, p. 129).” Some of these are explicitly aimed at language maintenance (prohibitions against bilingualism in either direction), and others which likely would have had a similar effect (maintaining the Arabo-Islamic character of the garrison cities.) It is not clear if these prohibitions would actually have been effective in maintaining the Arabic language, but they show that the early Muslim leaders were very concerned about maintaining their cultural and linguistic distinctiveness.

3.2.2.1 New Cities (*ʿamṣār*)

The desire to separate Arabic speakers from the mass of non-Arabic subjects may have been the impetus for the establishment of new, tabula rasa cities. These cities, called in Arabic *ʿamṣār* (singular. *miṣr*) were the primary form of settlement throughout the newly conquered territories, with the exception of the Levant.⁹ Whitcomb (1994, p. 161) describes three aspects of Islamic new cities

⁸ ولا نتشبه بهم في شيء من لباسهم في قلنسوة ولا عمامة ولا نعلين ولا فرق شعر ولا نتكلم بكلامهم ولا نكتني بكناهم (*ibid.*, p. 139)

⁹ There were at least two *ʿamṣār* in the heart of the Levant, but were unsuccessful as discussed in Section 3.4

(*amṣār*) that are important to the Arabization of new conquered territories:

- New cities were built on new land, even though they might be a short distance from existing settlements
- Conversion from temporary structures to permanent and sophisticated architecture was rapid and extensive
- Almost all these foundations were successful and endearing

The first aspect, building at sites removed from existing settlements, may be a result of the original function of these cities, which often began as siege sites. However, as we have seen there was a deliberate attempt to separate the conquering and subject populations, and this may explain why these siege cities were built up instead of abandoned for the conquered cities. More importantly, by being separate, these cities provide a linguistic and cultural beachhead for the largely Arabic speaking Muslims, who would not be dissolved into the general cultural and linguistic milieu that they found themselves in. Instead, they would promote bilingualism in the direction of Arabic, as nearby non-Arabic speaking peoples would likely have to travel to the new settlements to engage in trade with their new neighbors, where the dominant language of interaction would be Arabic. This could even lead to child language acquisition, as the children of the immigrants to these cities would interact with their peers and thus acquire Arabic early in life.

The second aspect, the conversion to permanent structures, suggests that these settlements weren't simply camp-sites, but rather (loosely) designed urban spaces, intended for permanent settlement. Again, this suggests that these linguistic beachheads were designed to last, and as per Whitcomb's third principle, they often did survive into the modern era — cities such as Fustāṭ, now Cairo, and Basra in Iraq are clear examples of this.

3.3 Arabian Peninsula

In this section, we will describe the history of what might be termed the Lesser Arabian Peninsula, those areas which are south of the modern Levantine states (Jordan, Israel, Syria, Lebanon) and Iraq.

The Arabian Peninsula is often seen as the heartland of Arabic speaking peoples, a view that only recently has begun to be challenged, as discussed further in Section 6.9. There were a number of languages spoken in the Arabian Peninsula prior to Islam: various Old South Arabian (OSA) languages were spoken in Yemen, and at the very south edge of the Peninsula, between Yemen and Oman, we find a small group of languages which are known as the Modern South Arabian (MSA) languages, with no direct relation to the OSA languages. We also find thousands of inscriptions in what are termed Ancient North Arabian (ANA) languages. These inscriptions are found throughout the Peninsula, but primarily near and in oases in the north Hijaz which had been minor city states. A number of inscriptions are also found in the desert near southern Syria, Jordan, and northern Saudi Arabia, apparently inscribed by nomads as opposed to the largely urban inscriptions around oases. The following section will deal specifically with ANA, and the other languages will be discussed with regards to their specific geographical region.

3.3.1 Ancient North Arabian and Arabic

The ANA languages are fairly diverse, and are attested beginning in eighth century BCE to the fourth century BCE, at which point they essentially disappear from the epigraphic record. These languages were used by “the settled peoples and nomads of central and north Arabia and by the nomads in what is now southern Syria and southern Jordan”, but probably these languages coexisted with old forms of Arabic, which would have been a primarily spoken language Macdonald (2008a,

pp. 180-181). There are numerous mixed texts which show features of both Arabic and ANA and possibly even OSA, and it is often, given the evidence, difficult to determine which ‘language’ these inscriptions are in (Macdonald, 2000, pp. 50-54). Such mixed texts suggest that the barriers between these languages were rather permeable and that multilingualism was probably relatively common.

The linguistic divisions between Arabic and ANA are relatively limited. Huehnergard (ALS) gives the a number of innovations which unite ANA and Arabic, suggesting that they shared an ancestor which branched off of Central Semitic. For this reason, I suggest we refer to ANA and Arabic as ARABOID languages, similar to how OSA and MSA are treated as language groups with a convenient title, in spite of their internal historical diversity.¹⁰ The features which are shared by the Araboid languages are:

- The *maḥḥu:l* form as a passive participle, but the ANA variety Dadanitic has MQTL and QTL for ‘killed’¹¹
- Merger of proto-Semitic *s and *š, though they are still distinct in one variety, Taymanitic

He also gives the following innovations in “proto-Arabic” which distinguish it from ANA where we have evidence:

1. No nunation (instead of mimation) in the case markings, but Arabic might show traces with words like *fam-un* ‘mouth’
2. Feminine demonstrative forms with *ta:*

¹⁰ I prefer the rather ugly ‘Araboid’ to Macdonald’s (2010) “North Arabian” since using a geographic designations has a tendency to bias our analysis.

¹¹ Classical Arabic has *qati:l* in the same meaning.

3. Relative *(a)llaði*:
4. Preposition *fi*:
5. Negative *ma*:
6. Verbal particle *qad* from *qadama*

However, many of these criteria are not true of all Arabic varieties, and therefore were probably not part of ‘proto-Arabic’: Item (1) has relics of its earlier form, and could have developed relatively recently in Arabic. The feminine demonstrative forms are certainly an innovation specific to Arabic (see 5.3.1.1) but only in certain dialects. The relative in (3) is only attested in a small minority of dialects, and deriving the widespread form of the relative *VllV* in modern dialects from that form is unnecessary — instead, it seems likely that the demonstrative forms *ði*, *tī*, *ði:n* were affixed to *VllV* in only a small minority of dialects (see Section 5.3.1.1) and a number show quite different forms entirely.

The preposition (4) is attested in some ANA inscriptions (Macdonald, 2000, fn. 205, noting that there is variation between common B and F in some Safaitic inscriptions). The negator (5) is widespread in Arabic dialects, but its use is more limited in Classical Arabic, which could suggest a development much later than “proto-Arabic”. Finally, the verbal particle *qad* is very limited in its distribution in modern dialects, and could also be a more recent innovation.¹²

¹² See also the discussion in Al-Jallad (2012, pp. 81-85) where he rejects most of these criteria (though on grounds that they could be ‘areal’ features) also but emphasizes the Arabic innovations (3) and (2), which as we discuss in this dissertation (in Chapter 5) reflect later processes of diversification than that of proto-Arabic. This leaves the conjugation of hollow verbs in ANA, and the dual pronoun with final Y (though the latter is only attested in a single ANA variety) to distinguish between ANA and Arabic.

Finally he lists innovative (with a following question mark in parentheses) features of ANA:

1. Pervasive assimilation of $nC > nn$ (shared with Northwest Semitic languages, Akkadian)
2. 3rd person dual suffix in Dadanitic written -HMY reflecting *-humay* versus Arabic *-huma*:
3. Some past-tense verb forms with a middle weak radical maintain the radical rather than replacing it with a long vowel, so ḤWR instead of verbs like Arabic *ka:na*

The assimilation of $nC > nn$ might not be innovative, as he himself admits, and is more widespread in Arabic than is generally acknowledged. This type of assimilation is well attested in Quranic recitations, where the final *-n* of the case markings regularly assimilate to following consonants, and dialects show various forms of n-assimilation (e.g. Syrian [wayrra:yiḥ] < *wayn ra:yiḥ*). It is possible that ANA simply generalized n-assimilation that was already present earlier on, and which Arabic dialects did not generalize nearly as widely. The second innovation is marginal, and the final one is also marginal, attested only in a minority of forms, so was either a change in progress or one being wiped out.

The primary distinguishing feature between the languages is the realization of the definite article, with *ha(n)* attested in ANA but *?al-* or *?am-* attested in Arabic, though Pat-El (2009) argues that they are from the same source. This form does seem to operate as a shibboleth, distinguishing the two groups, but many texts show such varying relationships between the definite article and other lin-

guistic features that Al-Jallad (2012, 183 ff.) treats this as an “areal feature” which spread from Arabic to ANA.¹³

Overall, the linguistic features differentiating ANA and Arabic are relatively minor, and many could have developed well after whatever supposed proto-language period where these languages “split”. They were likely mutually intelligible, as argued by Macdonald (2008b), one of the primary authorities on ANA:

It is difficult to believe that in practical terms there was a greater difference between [southern Arab tribes and ANA speakers] than there is between Syrian Bedouin and Sanʿānī Arabic today, i.e. a clear difference but not sufficient to impede mutual comprehension.

We also know that the languages must have been in relatively frequent contact. We know that at least from beginning of the current era, speakers of the two languages occupied similar territory, and that they have many shared common innovations. The many ambiguous inscriptions that show features of both languages suggest that features diffused regardless of language boundaries, that is to say ANA and Arabic speakers probably formed a single speech community which developed together after their initial ‘split’ on a historical linguists’ tree. The relative similarity of these languages would make it easier for changes to diffuse through both, causing them to become apparently more similar over time. A large tip in the balance of prestige from ANA languages to Arabic could cause sufficient diffusion to make it difficult if not impossible to tell if a language that now would be consider ‘Arabic’ had ANA roots.

¹³ Of course, in the approach taken here, every linguistic change is at its essence areal — the important criterion is when a form was adopted in a particular speech community. Indeed, the value of placing such mixed texts into one or another category seems to be limited, and tells us far less than the very fact that they are mixed.

We have some evidence that a cultural or historical event did indeed lead to the rise of Arabic at the expense of ANA languages. While most of the ANA Arabian languages which we have attested were used in settled oases, where each oasis adapted its own script for writing, there is a vast collection of graffiti clearly inscribed by nomads in the basalt deserts south-east of the Hawran, collectively known as Safaitic.¹⁴ The earliest dated example is from the mid-sixth century BCE, and the last dated example is from 267 CE, but it is assumed that this graffiti lasted into the fourth century before dying out (*ibid.*, p. 16).¹⁵ After this time, we have very few inscriptions, until the sixth century when we begin to find an increasing number of Arabic inscriptions, though not until fairly late do we find them in a specifically 'Arabic' script.

What seems to have happened to the writers of Safaitic is that a cultural shift took place in the northern Arabian peninsula. The Nabateans, who primarily used Aramaic for their inscriptions, were by the turn of the era probably largely Arabic speaking, but they had a division of functions: Aramaic was used for written activities, and Arabic was used for oral activities (*ibid.*, pp. 19-21). Other Arabic speakers were not necessarily accustomed to writing in Arabic, and do not seem to have had a cultural propensity for epigraphy, what Macdonald (2008b) calls the 'epigraphic habit.' A cultural shift, or even a replacement of ANA-speaking populations by Arabic-speaking populations, could have resulted in the loss of the epigraphic habit, with the ANA population possibly still existing but no longer recording themselves in the annals of history in the same way. This process appears to

¹⁴ Since the nomads would have had to learn the use of writing from settled people (Macdonald, 2010, pp. 15-16), this also provides us with evidence that there were friendly and fairly close relationships between the nomadic and sedentary users of ANA languages.

¹⁵ Though as Macdonald (*ibid.*, fn. 28) notes, the reason for the assumption that these inscriptions, which are rarely dated, end in the fourth century is that they make no reference to Christianity. In his words, "this is very unsatisfactory, but at present we have no other evidence."

have been complete in the fourth century, when the last Safaitic inscriptions are found, though they are primarily dated by their lack of reference to Christianity rather than on other bases.

As discussed below, a population replacement model is possible, since we have records of migration, though probably it was a combination of physical and cultural domination. These Arabic speakers would have largely displaced the ANA speakers (or cause them to shift further towards Arabic) and at the same time, cultural norms would have shifted to a more oral tradition (possibly one which explicitly valued oral transmission over writing).¹⁶ A situation of shift like this should leave some traces and we would expect to find some instances where ANA features are preserved in Arabic dialects. Most of these dialects, however, would be in the Arabian peninsula where our descriptive dialectology is rather poor, and so it seems that for present circumstances we can only really talk about the history of Arabic. It might be valuable, as more evidence becomes available for ANA and for peninsular Arabic, to develop a sense of how the Araboid languages developed together. However, for the analysis here, I will focus primarily on the Arabic varieties which gave rise to modern Arabic dialects, insofar as they are indistinguishable from ANA or ANA-Arabic hybrids.

3.3.2 Population movements in the Arabian Peninsula prior to Islam

As early as the first century CE, the incense trade, the source of income and power for the South Arabian kingdoms, was shifting northward, or into the Red Sea instead of overland. The region retained enough importance to be the site

¹⁶ This could be interpreted, as Knauf (2011) does, as meaning that ANA was the ancestor of Arabic, but that does not hold in the sense used by historical linguists, as explained by Al-Jallad (2012, pp. 83-5). However, it does mean that there were links far more intimate than those of common descent between the languages of these speech communities.

of a power struggle between the Roman and Persian empires, in spite of the distance between those empires and Southwest Arabia. The battles over this territory by foreign powers helped propel the decline of the Old South Arabian kingdoms: Roman invasions into Yemen in the first century BCE, the proxy wars of Aksum against the Sabeans and Himyar in the fourth and sixth centuries CE, the final Persian invasion in the sixth century prior to the Islamic conquests. Roman and Persian powers also established and financed important frontier client states in the Levant and Iraq, further shifting power to the north. In the early sixth century, the mining of precious metals in the central Hijaz gave the residents of that area more access to luxury goods from elsewhere, increasing the importance of trade and trade cities like Mecca (Kennedy, 2007, p. 44). The overall picture, then, seem to be the growing importance of central Arabia, while southwestern Arabia, though still strong, was becoming relatively weaker.

These factors appear to have led to a number of populations movements out of the south-central peninsula into other regions. The Ancient North Arabian (ANA) inscriptions start to give way to more clearly Arabic inscriptions (Hoyland, 2009, p. 391) in the third or fourth century CE as discussed above. Moreover, tribe names recorded in ANA inscriptions appear to be largely absent from Muslim geneologists' lists of tribes (*ibid.*, pp. 384-385). Tribes that do appear in both the inscriptions and the geneological tradition are first mentioned in central and south Arabian inscriptions (usually in OSA languages) but later show up further north. For example, the tribe of Ġassān is recorded in central Arabia in inscriptions dateable to between 260–360 CE, but they are well entrenched in the Hawran as the Ghassanid client state of the Romans in the sixth century (*ibid.*, p. 386).

Accounts of northward movement are also found in the indigenous Arabic historiographical tradition, where tribes are, for various reasons, forced to leave

their ancestral homelands in the south and move north. One of the most intriguing of these accounts (others quoted in translation in Hoyland, 2009, p. 388) is that from the 3rd/9th century historian Al-ʿAṣmaʿī's *tārīḥ al-ʿarab qabl al-ʾislām* 'The History of Arabs Before Islam' quoted by Hoyland (*ibid.*, p. 388):

They (the southern 'Arab' tribes) did not enter a land without robbing its people of it. Khuzā'a wrested Mecca from Jurhum; Aws and Khazraj wrested Medina from the Jews; the clan of Mundir seized Iraq from its people; the clan of Jafna seized Syria from its people and ruled it;¹⁷ and the progeny of ʿImrān ibn ʿAmr ibn ʿĀmir [of al-Azd] seized Oman from its people. Up till then all of these [southern tribes] had been in obedience to the kings of Himyar.¹⁸

The Hijaz, perhaps the original center of the Nabatean trade empire (Knauf, 2011, p. 22), was growing in importance and tribal groups seem to have been moving up and through that corridor. There was an increase in economic activity at this time in the Levant, which could have been either the cause of these migrations, or the result of them, or possibly both. In either case, that would indicate an increase in population in that area (Hoyland, 2009, pp. 387-8). The trading empire centered in Palmyra had been sacked by the Romans in 272, which may have created something of a power vacuum that encouraged migration into the area, and this is also a part of the Arabic mythology (*ibid.*, pp. 389-90).

¹⁷ The last two would be the Lakhmid's and Ghassanids respectively, using the terminology of this dissertation.

¹⁸ This intriguing account fits well with C. Robin's (1991) fifth phase of Arab-South Arabian relations (see Section 3.3.3), where Arab groups were under the control of the Himyarite state, which itself was slowly breaking down.

Thus, it seems that speakers of languages which gave birth to the Arabic dialects began to spread northward from the central Arabian peninsula, particularly the central and south-central western edges of the peninsula, the Hijaz and northern Yemen, primarily in the third to fourth centuries CE. Iraq may also have been populated from further east, from the coastal area opposite modern Bahrain, called in the sources Bahrayn.¹⁹ There were east-west trade routes through Najd and Yamama to Bahrayn and there were close relations between those areas and Mecca, which bought grain from the eastern central peninsula, so it was natural for migrations to move in this direction. The Bakr b. Wā'il tribe that occupied the Yamama and the eastern coast shortly before Islam, for example, is said to have been part of the Kinda empire that was centered on Qaryat al-Faw, again suggesting a SW peninsular origin for even those tribes (Caskel, *EI2*, "*Bakr b. Wā'il*").

Hoyland (2009, pp. 386-7) prefers not to see this as a population replacement per se, in spite of the evidence of tribal names disappearing. Tribal names could have disappeared as a result of the coalescence of new tribal confederations, named after the newcomers, which might have subsumed previous confederations as they developed into unified political entities which could benefit from the patronage of the major powers of Rome and Persia (*ibid.*, p. 394). For example, the tribes of Nizār and Ma'add are treated as separate entities in the fourth century Namara inscription, but are merged in the Islamic genealogies (*ibid.*, n. 81).²⁰ One can also appeal to the fact that nomadic tribes particularly have very low population density, so that the "repopulation" of an area might be difficult to distinguish even from a seasonal migration, as described in Section 2.5.

¹⁹ The same pronunciation, but the alternate spelling is used in English language sources to distinguish the two.

²⁰ However, Al-Jallad (2012, pp. 94-109) has a different reading for this line.

3.3.3 Yemen

There were also important, and probably somewhat earlier, migrations southwards. The southern edge of the Arabian peninsula was, and in places still is, populated by speakers of languages other than Arabic, and the southwestern corner of the peninsula, what is now Yemen, was the home of a series of major world empires. However, with the shift of dominance northward, the Yemen kingdoms lost their power, and once the Islamic conquests shifted the centers of power even further north, in Damascus and then Baghdad, Yemen became a neglected backwater.

The oldest empires centered in Yemen date to the 12th century BCE, and continued to exist until almost the dawn of Islam. These empires all wrote in languages referred to as Old South Arabian (OSA) languages, Western Semitic languages in the Central Semitic branch and thus a sister to the Arabic branch in the tree suggested by Huehnergard (2005). Most of these empires derived their wealth from the incense trade, as they controlled both the sources of incense and the major trade routes that went through the Red Sea or along the coast of the Arabian peninsula. Their power was often far-reaching, being able to launch raids as far as the Persian Gulf in the east, or the Levant in the north.

The grip of OSA speaking peoples on this region appears to have been challenged from the north by Arab invaders beginning in the last two centuries BCE, while global politics in the first five centuries of the Common Era disrupted the rule of native kings in Yemen. At the same time, repeated breaches of the Marib dam, one of the most impressive hydrological works of the ancient world, combined with environmental degradation, compromised the vital Jawf area that had been the seat of power for the South Arabian kingdoms.

Arabs (*ʿrb*) are referred to in OSA documents as early as the sixth century

BCE, where the term seems to be used to refer to pastoralists, as opposed to sedentary agriculturalists, though no more attestations are found until the first century BCE (C. Robin, 1991, pp. 72-3). C. Robin (*ibid.*) gives five phases of Arab penetration into Yemen:

1. The occupation of the Jawf by Arabs (2nd to 1st centuries BCE)
2. Confrontation between Arabs and the Sabeans (1st-2nd centuries CE)
3. Sabean domination of the Arabs (3rd century CE)
4. Integration of the Arabs by the Himyarites (4th-6th centuries CE)
5. Seizure of power by Arabs (to the end of the 7th century CE)

The first phase is the incursion of Arabs into Yemen, particularly into the north-eastern Jawf region. This arid region, north and east of the highlands, and opening out into the major deserts of the peninsula, has very poor rainfall, but fertile clay soils. In antiquity, the Marib dam allowed major agricultural exploitation of this area which was the center of power of many of these South Arabian kingdoms. During the 2nd and 1st centuries BCE, a semi-independent kingdom of Haram arose, using a different language, pseudo-Sabaeen, which may have been OSA with Old Arabic interference according to C. Robin (*ibid.*) and E.Kogan and V.Korotayev (1997, pp. 221, 237-9). This kingdom, situated in the northern Jawf, may have been ethnically and even linguistically Arab, and represented the first major Arab power in the region. Around the same time, incursions by both Romans and Arabs led to a general reduction in the urban density of the Jawf (Schiettecatte, 2010), starting the area on the long route to decline.

The second phase consisted of a number of confrontations between the Sabean state and the Arab nomads, which were not particularly conclusive at that time.

The third phase consisted of more vigorous Sabean efforts to rein in the Arabs, with varying amounts of success. They invaded Qaryat al-Faw, at the border of the Hijaz and modern Yemen, then probably a major center of Arab (Araboid-speaking) rule. At other times, the relationship between the Sabeans and the Arabs came closer to a symbiotic one, with both groups benefiting from trade and many Arab groups becoming clients or protectorates of the Sabeans. However, during the Abyssinian reduction of the Tihama coast up to Najran, Arab tribes tended to side with the Abyssinians against the Sabeans, which may have increased Arab penetration into the Tihama (C. Robin, 1991, p. 80)

The fourth phase, beginning at the end of the 3rd century, witnessed the rise of Himyar, a major power which overtook the states of the Sabeans and Hadramawt, and which also extended its power over the Arabs of the southern peninsula. The Arab tribe of Kinda appears to have controlled areas outside of Himyar's dominions, basing themselves as usual at Qaryat al-Faw. The Himyarites integrated the neighboring Arab tribes into their empire, using them as auxiliaries, in a pattern similar to that of the Roman and Persian empires. Arabs seem to have spread more widely within Yemen at this time, with one inscription from the beginning of the fifth century referring to the Arabs of the Tihama plain and the mountains, apparently referring to 'Asir, the mountainous region immediately north of modern Yemen (*ibid.*, p. 81).

It was also during this period that Yemen was occupied by forces coming from Ethiopia, which likely undermined the Himyarite control of the region. First, from 340–378 CE, the Aksumite kingdom invaded and occupied the coastal littoral.

Some two centuries later, another invasion from Ethiopia resulted in subjugation of Himyarite suzerainty to Aksum in Ethiopia, which lasted from about 517 CE until 570 CE, when Sassanian troops invaded and placed Yemen under Persian control.

The final phase, beginning well before the Islamic conquests, seems to have witnessed an expansion of Arab dominance in the south-west peninsula. The conquests themselves solidified that dominance. Following this, Yemen was under the sway of a variety of Islamic dynasties.

By the tenth century, in spite of nearly two millennia of rule by Old South Arabian languages, the country was largely Arabized, with Arabization generally correlating with altitude, as seen in Figure 3.3. Lower lying areas seem to have been Arabized earlier, and bilingualism (referred in the diagram to ‘Arabic mixed with Himyaritic’) seems to have been the first step towards Arabization. Even today, some areas speak dialects which are difficult to classify between Arabic and OSA, as reported in Watson et al. (2006a).

Yemen has been, and continues to be, largely split along a NW-SE axis. The NE half of the country has less water, and with the reduced agricultural viability of the Jawf, those in the north pursue marginal subsistence agriculture lifestyles (see for an overview of this region see Dresch, 1989). In the southwest, which includes the Tihama coastal region, rainwater fed agriculture is much more successful, which leads to conglomerated, large estates often owned by absentee landlords and worked by laborers or sharecroppers. The two halves of the country have tended to be divided politically, with Shiite groups controlling the north, and Sunni groups in the south (*ibid.*, p. 11). This split can often be seen in the Arabic dialects of the country, so that, for example, dialects of the Tihama tend to lack the *ha:-*prefix in demonstratives, which is found in the north-eastern dialects.

The Arabization of Yemen, then, likely began with the penetration of Arabs into the northeast of the country as early as the second century BCE. It is unclear when Arabs began moving into the Tihama plains, but we have references to them being in that area by the third century CE, perhaps moving into the vacuum created by Aksumite invasions of that area. By the tenth century, Arabization had proceeded quite far considering the extremely difficult terrain of the area, with peaks as high as 3,500 meters, and an average elevation of 2000 meters, and the long tradition of another language family in the area. The Arabization process seems to have included long periods of bilingualism, with more elevated areas taking longer to shift to Arabic.

3.3.4 Oman

The movement of Arabs into the Jawf made them susceptible to problems with the Marib dam, which collapsed a number of times in its lengthy history. A collapse in the second century CE is said to have led to the migration of the Azd tribe into Oman (among other areas in the peninsula). Though at the time there were extensive Persian settlements along the coast, the Azd eventually was able to defeat them and settle the northern coastal littoral (Al-Rawas, 2000, p. 20). Further tribal migrations continued into the area, though many seem to have come from east or north-east of the Yemeni mountains.

There still remain, primarily in the large, arid region between the Yemeni mountains and the Omani coast, a number of languages referred to as Modern South Arabian (MSA) languages, which are more closely related to Ethiopic than they are to either Arabic or the Old South Arabian languages. These areas are largely shifting to Arabic, and many of these MSA languages face extinction. Nonetheless, there may have been and may continue to be some influence from these lan-

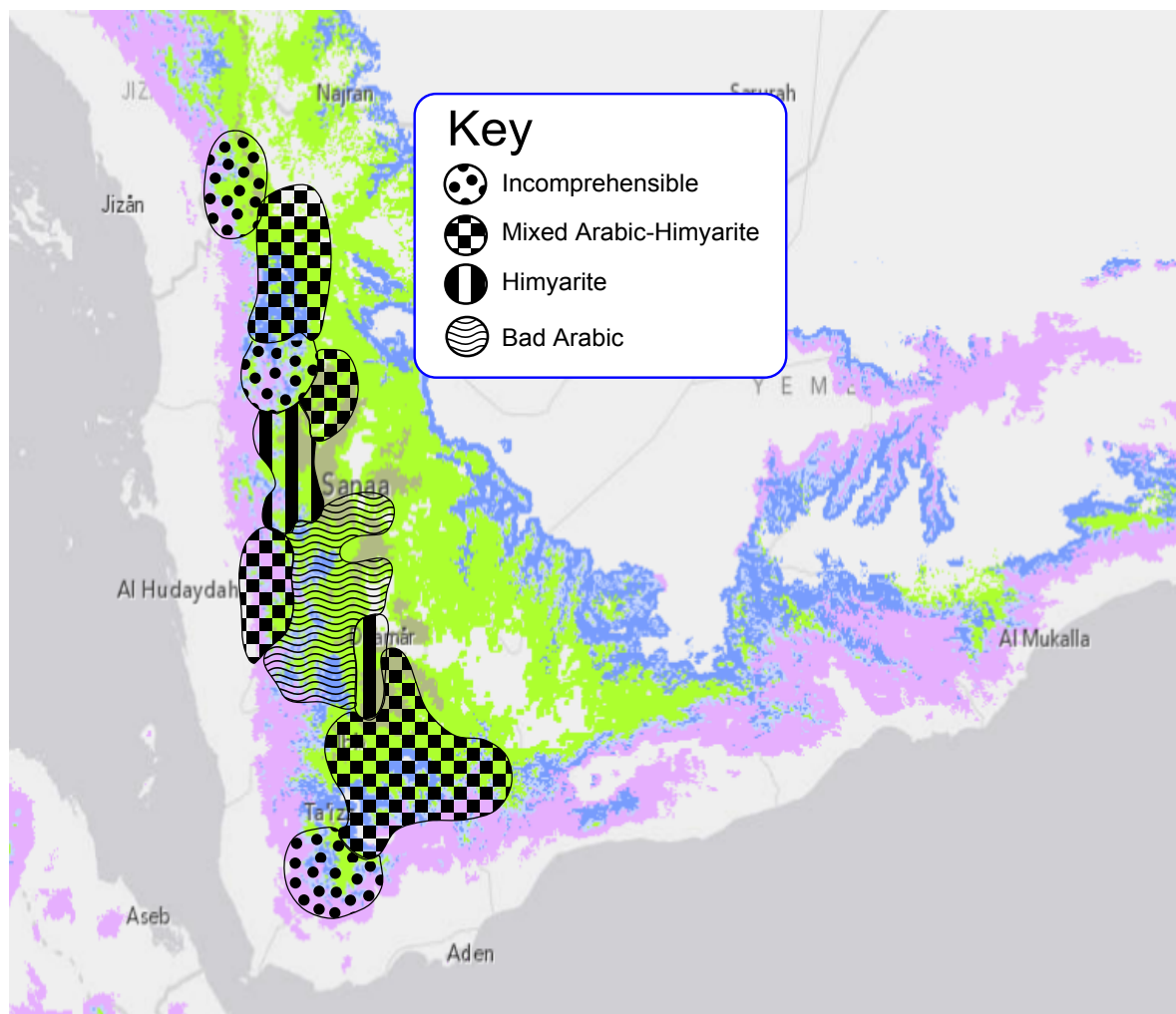


Figure 3.3: Map of speech in Yemen according to al-Hamadhani in the tenth century CE, adapted from C. Robin (1991), overlaid on a map of Yemeni topography. Areas not labeled are said to speak good Arabic.

guages on the Arabic dialects of the region.

3.3.5 Summary: Arabian Peninsula

The overall picture, then, is that prior to the Islamic expansions of the seventh century, two related languages co-existed in the Arabian Peninsula: Arabic, and Ancient North Arabian, what we call here Araboid languages. ANA speakers have apparently disappeared, possibly due to cultural or linguistic shifts resulting from a major diaspora of speakers of Old Arabic originating in many of the same places as the later Islamic expansions — the Hijaz and Yemen — who moved out into the northern and eastern parts of the peninsula, beginning as early as the second century. At the same time, gradual migration seems to have moved from south-central Arabian further south into Yemen as well, and various events such as the breaching of the Marib dam led the Arabic speakers living in those regions to move outward from the west-central Arabian homelands into places like Oman.

3.4 The Levant

Geographically, the Levant is a varied region, blending in smoothly with northern Arabia, with low-rainfall deserts extending almost to the Jordan river at the western extreme, and being bound in the north by the Euphrates and later the mountain ranges of Turkey. The anti-Lebanon mountains create a rain shadow, making the coasts lush and verdant at the expense of areas around Damascus on the other side of the range, while central Syria in the Orontes valley is relatively better watered. The deserts of Jordan, Syria, and Iraq are essentially contiguous, so camel herding nomads can move relatively freely between the Levantine and Euphrates river valleys, while small-stock nomads are more restricted to the steppe lands on both sides of the desert and are somewhat more bifurcated into Syrian

and Iraqi (using the modern country names) groups, with connections occurring further north, where the Euphrates comes close to Aleppo and rainfall is generally higher. See the map in Figure 3.2 for an idea of how rainfall in the Levant would influence settlement patterns.

3.5 Pre-Islamic Linguistic Situation

The Levant was largely the domain of Aramaic speakers prior to the Islamic conquests. Starting in the first millennium BCE, Aramaic was both the spoken language and the primary language of writing and administration in the Levant. With the arrival of the Seleucid Empire in the third and second centuries BCE, Greek became the primary language of administration in the area, a tradition maintained by the Roman, then Byzantine empires. Nonetheless, literary texts (including large parts of the Bible) continued to be composed in Aramaic, and it was the primary administrative language of the Nabatean empire (ca. 168 BCE to ca. 106 AD, at which point it was annexed to the Roman empire, but continued to exist), which later seems to have shifted largely to speaking Arabic, but largely writing in Aramaic.

Greek had little success in becoming the spoken language of the Levant, in spite of nearly nine hundred years of Greek administration prior its replacement by Arabic in the seventh century CE. In Madaba (modern central Jordan) we have a sixth century inscription in Greek that quotes the locals speaking and making reference to the bible in Aramaic (Hoyland, 2004, p. 187). A pilgrim from the 380s AD in the same area is quoted as saying:

In this province there are some people who know both Greek and Aramaic; but others know only one or the other language. The bishop may

know Aramaic, but he never uses it. He always speaks in Greek and has a presbyter besides him who translates the Greek into Aramaic so that everyone can understand what he means. (Hoyland, 2004, p. 187)

Moreover, while Madaba and other Aramaic-speaking urban areas wrote in Greek, smaller villages in the countryside used Aramaic in inscriptions. Aramaic, specifically Syriac, was also used in northern Syria and Mesopotamia as a major vehicle for inscriptions and literary composition. This is probably a result of the comparatively lighter Hellenization of those northern areas, since Roman rule came nearly 300 years later than in southern Syria, Palestine and Jordan (*ibid.*, pp. 187-189).

Thus, even as close to the steppe as Madaba, east of the Jordan river, in an island of arable land surrounded by more marginal lands, Aramaic seems to have been the primary spoken language, with Greek operating as a primary language of culture. It is likely that a similar situation was true of areas north and west of here, with Aramaic as the primary spoken language.

This supports our theory that those who claimed or had ascribed to them an 'Arab' ethnic identity, but settled in largely Aramaic speaking areas, likely shifted to Aramaic. Macdonald (2009b, VI, pp. 313–314) cites a source from the third century BCE as call areas even further north in Jordan, including Amman, 'the territory of the Arabs', but it would seem that by the fourth century CE, Aramaic was the dominant language in that same area, even if it was hidden at times beneath the Greek used as an inscriptional language.

Immediately beyond the limits of arable land, in the steppe and the desert, the primary languages were Araboid. In the Nabatean kingdom, situated around the 'saddle' that straddles the Sinai-southern Jordan, and the north-west corner of

the Arabian peninsula, and thus largely in steppe lands, there was at least some Aramaic-Arabic bilingualism as early as the turn of the era from BCE to CE. Legal documents discovered as part of the Dead Sea scrolls show a large number of Arabic loan words, suggesting a “sophisticated cultural role” for Arabic among the Nabateans at this early date (Macdonald, 2006, p. 467). A later report from around 370 CE claims that Nabateans in both Petra and the Negev desert (modern Israel) sang hymns in Arabic, though it is not entirely clear what is meant by this (or how the author could distinguish between Arabic and Aramaic.)

One of the earliest inscriptions in something resembling Arabic is JSNab 17, a funerary inscription from al-Hijr, now Madā'in Šālīḥ in NW Saudi Arabia, that is dated to 267 CE. This text appears to be attempting to use Aramaic, but filling in “gaps in [the writer’s] knowledge with Arabic words and phrases”, suggesting an unequal bilingualism with Arabic dominant (*ibid.*, p. 471).

Another inscription of note is the Zebed inscription, part of a church dedication lintel that has three languages on it — Greek, Aramaic and Arabic — though it is not tri-lingual per se in that each language expresses different content (the Arabic being little more than a list of names). Zebed is quite far north, only about 40 miles SE of Aleppo, but it is located in what is clearly steppe lands.

3.5.1 Client States and the Arabization of the South-central Levant

The earliest Arabization outside of the stepped was probably in the Hawran, a largely arable basalt region south of Damascus, stretching from the Golan Heights in the west to the desert steppe in the east. The Hawran is a natural transition point for movement from a nomadic to a settled lifestyle. It is a largely arable basalt region south of Damascus, stretching from the Golan heights in the west to the desert steppe in the east. This area is extremely fertile, as a result of the volcanic

soils underlying it, and receives sufficient rainfall for catchment agriculture (and indeed, has no major waterways.) However, in the east it simply transitions to steppe lands, and so the Hawran as a region likely showed the full spectrum of settlement, from agriculturalists in the west to small-stock nomads in the east. Major trade routes, such as the Wādī Sirḥān (Donzel, *EI2*, “*Sirḥān*”) lead from the Hawran south-east into the deserts, thus bridging the fertile ‘interior’ with its major urban centers and the deserts that trade flowed across.

The Arabization of this area was made possible by an important change in Roman military policy. Beginning in the second century CE, Roman policy focused on establishing frontier fortresses, the *limae*, along the edges of the desert with the mission of curtailing raids from pastoralists (Mayerson, 1986). These fortresses evolved to serve a wider variety of purposes, including trade and administration (cf. G. Fischer, 2004, p. 54), and may themselves have been responsible for a limited amount of peaceful contacts between Araboid nomads and Aramaic and Greek speaking settled peoples. We would expect to see one-way bilingualism in this situation, with Arabic speakers learning Aramaic or Greek for trading purposes.

However, military and financial pressures eventually forced the Romans to abandon these forts and adopt a different strategy.²¹ Instead of maintaining an expensive (in cost and manpower) frontier against what was only a dubiously

²¹ The *limae* were completely demobilized by order of Justinian in 551, though many of these garrisons appeared to have slowly been abandoned starting as early as the fourth century CE (*ibid.*). This freed up their manpower for other theaters, such as Western Europe and the areas around the Caspian. G. Fischer (*ibid.*, p. 51) notes that in 395 CE, the number of troops deployed on the Arabian frontier was approximately equal to the number of troops deployed along the Persian border, so reshuffling those troops to the more vital Persian border would have been a significant increase in the troops available to deal the Sassanians, the other major world power whose political power was focused in Iraq, though they were culturally and linguistically Iranian. Moreover, the Byzantines appear to have had a somewhat exaggerated concern with Persia, often disproportionate to the actual threat they posed (*ibid.*, p. 53).

dangerous threat, they began to outsource the frontier security to Arabic speaking tribal federations. They chose a dominant tribe in the area to police the other tribes in the area and in times of war to fight as auxiliaries in the Roman armies. In return, these confederates were allowed to settle on Roman lands, and were given various economic and political incentives to tie them to Roman power, as well as to help them maintain their ascendancy over other tribes.²²

From a linguistic perspective, these *foederati* of the Romans, settling in the Hawran region, would have been the first major Arabic linguistic colony in the arable lands of the Levant.

The earliest Roman confederate in the region, the Arab tribe of Ṣalīḥ (Shahîd, *EI2*, “*Salīḥ*”), appears to have originated in the Tihama coastal plain of the Arabian peninsula, but moved into the Arabian side of the Wādī Sirḥān, and likely settled near the Hawran, where they became Roman clients ca 400 CE (Shahîd, 1989, pp. 244-6). A century later or so, the Ghassanids moved into the area, settling within Roman territory in 490 CE, and overthrowing Ṣalīḥ shortly thereafter, becoming the new clients of Rome in 502 CE (Shahîd, *EI2*, “*Ghassān*”, 1989, pp. 282-3).

Even earlier than this, however, the Namāra inscription dated to 328 CE , boasts that the entombed king Mr’ l- Qys’s sons became “chief men of Rome” (Al-Jallad, 2012, p. 110). This inscription, a grave stella written in Arabic (in Nabatean script) is located in the eponymous town approximately 60 miles SW of Damascus. It is one of the longest and earliest inscriptions in a language that is recognizably Arabic, though the deficient nature of the script leaves it open to a number of

²² We will not consider here the political or historical import of these arrangements, since the focus here is primarily linguistic in nature, but see Donner (1981); G. Fischer (2011); Hoyland (2001); Shahîd (1995, *EI2*, “*Salīḥ*”).

different interpretations (for a comprehensive overview of the inscription, see Al-Jallad, 2012, pp. 94-113).

However, for the purposes here, the Namāra inscription indicates that as early as the fourth century CE, Arabic speaking client states of the Romans had taken up residence in the Hawran. Moreover, the highly unusual step taken by Mr' l- Qys of having his grave inscription written in Arabic suggests that this represents a “conscious and deliberate choice [...]” presumably “to make a statement about [his] ethnic and/or cultural affiliation, about [his] Arab identity. (Hoyland, 2004, p. 184).” Thus, not only is there a presence of Arabic in the region by this time, but there is also a growing desire to assert the value of the language over its rivals, Aramaic and Greek, by using it in this important inscription.²³ There are a number of other, later inscriptions throughout the Hawran, and they are in an increasingly recognizably “Arabic” script. Indeed, some of the greatest concentrations of clearly Arabic inscriptions are in the this area.²⁴

Christianity and Christianization presented another motivation for settlement of nomadic peoples in the area, since a settled lifestyle was often seen as a necessary part of conversion to Christianity. The Roman ideological biases against nomads (Shaw, 1982) influenced Christian thought as well, and thus “it is very likely that settlement pressures were brought to bear on steppe Arabs as they were introduced to Christianity” (G. Fischer, 2011, p. 42). Conversion to Christianity was extremely important for the Arab allies of Rome, and may have even been a pre-requisite for operating as a *fordoerati* (G. Fischer, 2004, p. 55).

²³ One might add to this the various epigraphic languages which constituted ANA.

²⁴ We cannot completely ignore the possibility that this distribution is the result of modern political circumstances, which allow archaeologists easy access to Israel and Jordan, but relatively less to Syria, Lebanon and Iraq, as much as the actual distribution of inscriptions. However, this is well beyond the scope of this paper.

The Arabization of the Hawran therefore is a direct result of Roman policies to support Arabic speaking client states settled Roman lands. Indeed, the outsourcing of both defense and eventually taxation to these client states seem to have led to an era of prosperity in the region, which likely encouraged greater sedentarization. Remains of settlements in the area show few fortifications, in contrast to settlements further to the north which had to expect the possibility of Sassanian attacks (G. Fischer, 2011, pp. 103–104), making the Hawran one of the more attractive places to settle down in this era. This prosperity, combined with the impetus to convert to Christianity and settle, likely drew Arabic speaking nomadic groups yet further into the heart of the Levant. Finally, the political power enjoyed by these *fodoerati* also may have increased the prestige of the Arabic language itself, as reflected in the somewhat audacious choice of Arabic as the language of inscription in the tomb of MR' QYS.²⁵

3.5.2 From Roman to Arab Rule

In contrast to the policies of the Romans, who themselves had little stake in the Arabization of the Levant, the policies of the later Arab conquerors did relatively little to establish viable linguistic colonies in the Levant. Instead, as we will see, it was likely a series of natural disasters that led to an increase of Arabic speakers in the region.

However, before the Arab conquests began, another major political event influenced the linguistic scene in the Levant. The Sassanians invaded the Levant in 611, and were not evicted until 630, an event that may have fundamentally dis-

²⁵ However, the following client states tended to use other languages for inscription, especially Greek. Inscriptions from the Ghassanid's are primarily in Greek, with some exceptions such as the Harrān inscription from 568 CE which is bilingual in both Arabic and Greek.

rupted what had been a nearly five hundred year period of Roman domination of the Levant. It may also have reduced some of the political and linguistic resistance to the incoming Arabic speaking Muslims. For the first time in centuries, potentially an entire generation would have been raised that only knew Sassanian rule (Donner, 1981, pp. 99-100). Tribal alliances that had been fostered by the Byzantines likely dissolved, potentially prompting new tribes to move into areas they previously had not been able to occupy due to Ghassanid resistance. Even after they reasserted their power over the Levant, the Byzantines were unable to control as much territory as before, with their southernmost control reaching only the Dead Sea rather than Ayla (modern Aqaba) as it had before (*ibid.*, p. 100). Thus, the Arab conquests essentially continued the de-Hellenization of the Levant that had been begun by the Sassanians.

The most immediate consequence of these invasions, beginning in 12/633²⁶ and ending in 27/648, would have been the emigration of most of those with the strongest ties to the Byzantines. These groups would have had the strongest connection to Greek as a language of administration, learning and the Greek Orthodox Church. Aleppo is said to have seemed abandoned when the Islamic armies arrived, as many of the inhabitants had already fled to Antioch, and other cities were said to have been similarly abandoned (*ibid.*, pp. 245-246). Indeed, the Byzantine policy after these military defeats was to evacuate rather than defend its citizens. When Tripoli was besieged, the ships sent by Constantinople were not reinforcements; rather they were sent to evacuate the populace, leaving the town to the Muslim conquerors.

Those bulk of those who remained were Aramaic speakers, and their con-

²⁶ For dates in the Islamic era, I will follow the scholarly convention of indicating first the Islamic data, in lunar years, following by the standard Gregorian solar calendar

nection to the Byzantine Empire, and Greek as an important language, was probably tenuous. First, there was the disruption in continuity due to the Sassanian invasion from 611–630, followed by the brief period of reassertion of Byzantine rule, which likely had already reduced the allegiances of many of the people in the Levant.

Second, there was the ongoing struggle between the Byzantine orthodoxy and the Monophysite Christianity practiced in the Levant. This split in the loyalties of Levantine residents is reflected in a larger ideological-linguistic split, such that “in Greek writings, Muslims are only ever the enemies of God”, while in Aramaic writings arguments against Islam are expressed in more measured terms (Hoyland, 2004, p. 194). Later writers even viewed the Islamic conquests in rather rosy terms. Dionysius of Tellmahre (d. 845) states that “If, as is true, we have suffered some harm [from the Islamic conquests] nonetheless it was no slight advantage for us to be delivered from the cruelty of the Byzantines (*ibid.*, p. 103)”, though accounts at the time the conquests were not nearly as optimistic.

The immediate linguistic result of the conquests, therefore, was to further attenuate the links between Levantine Aramaic speakers and the Greek of the Byzantine empire, and perhaps to seriously reduce the number of fluent Greek speakers in the region.

3.5.3 Urban settlement

On the other hand, the new conquerers did little to establish viable linguistic colonies in the Levant, in contrast to many of the other regions they conquered. Unlike in Iraq and Egypt, the conquerers of the Levant did not found very many *amṣār*, ex-nihilo cities which formed a linguistic beachhead in other regions (see section 3.2.2 for more on these). There were of course garrison cities, but few of

them seemed to have been in areas where Arabic was a minority language. Ayla, at the site of modern Aqaba, was obviously in the heart of the Arabic speaking, Nabatean homelands where the Arabian peninsula blends into the Levant. Ayla and similar settlements, such as the Umayyad desert castle complexes, were designed to control the desert dwelling nomads and desert trading routes that became increasingly important in the Islamic era, but did not operate as linguistic colonies since they were already in Arabic speaking areas.

Two other major cities, al-Jābiya in the Golan region (formerly the capital of the Ghassanids) and Ramla in NE Palestine, were founded as largely Muslim settlements, but they appear to have failed to survive, let alone thrived. The 'Amwās plague of approximately 18/639 likely did away with these cities, reducing the population of al-Jābiya from 24,000 to 4,000 (Donner, 1981, p. 245). This city was in an already largely Arabic speaking region and therefore even if it had survived it would not have played a large role in Arabizing the Levant.

The exception to this rule were Umayyad garrisons along the coast in what was firmly Aramaic and Greek speaking territory (El'Ad, 1982). These garrisons were originally basic military colonies, but they were expanded into self-supporting cities similar to the *amṣār*, with the holders of the garrisons granted land to work. They were initially settled with Arabic speaking soldiers, but even with incentives of land and pay raises, they were not able to find sufficient Arabic troops, and so they were also settled by non-Arabic speaking troops, primarily Persians, though these were Persian converts who had initially moved into the area during the Sassanian conquests and lived in areas around Baalbek and Homs. However, these coastal cities were repeatedly destroyed, and may have been reinforced with troops from further inland. Harris (2012, p. 35) even reports that after the fall of Tripoli, the Muslims were so desperate to bring the city back up to a functioning level, they

populated it with Jewish settlers. Therefore, they may have gained significant populations of Arabic speakers at some time, though the evidence is not clear, and we cannot be sure that these coastal cities formed viable Arabic linguistic colonies.

Instead of creating new cities, the victorious conquerers chose instead to move into intact, but partially abandoned cities. Many of the treaties that the Muslim armies signed with conquered cities stipulated that a certain percentage of the dwellings in a city be handed over for use by the conquerers, and it is likely that this was not too onerous for those already living in those cities, since many of the Byzantine residents would have fled to Anatolia. In Homs, sources indicate that the city was divided among the Muslims and they were “settled in [...] every place whose occupant had evacuated it and in every abandoned yard (Donner, 1981, p. 247)”. The original inhabitants still living there seem to have moved further away, settling along the Orontes river (*ibid.*, p. 247). Similar accounts exist for other cities in the region, suggesting a general trend of repopulating the abandoned cities with Arabic speakers.

The Arabic speaking armies also settled in a limited way in the countryside. As in Iraq, the Islamic state of this time discouraged the dispossession of already resident and productive peasants, so any settlement had to occur on lands which had been abandoned or which had never been cultivated in the first place. Restoring (and maintaining) the agricultural yields of the newly conquered lands was extremely important to the new conquerers, and grants of land often stipulated that the new owners restore them to productivity (*ibid.*, p. 248).

Who then were the groups that settled Syria, and where were they from geographically? First, it is important to note that the total quantity of immigration and settlement seems to have been limited. Unlike the new cities of Iraq and Egypt, which were constantly absorbing new waves of settlers, *rawādif*, we have almost

no mentions of in-migrating waves in Syria (Donner, 1981, p. 249). Donner (*ibid.*, pp. 250-251) argues instead that Syria was likely settled largely by the warriors who conquered it, the same warriors who made stipulations about providing housing in major cities. Moreover, the tribe of Quraysh, especially the close compatriots to Muhammad, his successors, appeared to view Syria, with which they had close trade links, as their personal preserve, hence the choice of Syria as the base for the first Islamic dynasty, the Umayyads.

The initial settlers of Syria were largely from the Hijaz, but were also supplemented with soldiers from “Yemeni” tribes, both lowland and highland tribes, with tribes from both northern Yemen and the Hadramawt participating. The forces, therefore, were predominantly from the L-shaped area of the south-west Arabian peninsula. However, all things considered, the numbers of soldiers were relatively small — estimates of the Arabic force at the battle of Yarmouk are around 24,000. This figure may not include extended family that the soldiers brought with them (the women were said to have been called upon to fight in the battle at one desperate point), but combined with the report of the population of al-Jābiya above, suggests total numbers of in-migrating Arabic speakers in the low tens of thousands.

The evidence thus suggests the settlement of the Levant was rather limited even into the Islamic period, which makes it difficult to account for the eventual Arabization of the Levant. Though there was a beachhead of Arabic speakers in the Hawran, they and the Arabic speaking Muslim conquerors would have still been a tiny minority, in spite of their political dominance over Syria. Even in cities that were abandoned by the Hellenized Byzantine elites, the vast majority of the population would have been Aramaic speaking Christians. Thus, the situation was not, as one author put it, of the *Church in the Shadow of the Mosque* (Griffith,

2008), but rather quite the opposite, with mosques firmly in the shadow of the surrounding churches (Tannous, 2010). All things being equal, these settlements would not have been viable linguistic colonies, and should have been absorbed in a melting-pot effect, with the minority of Arabic speakers dissolving into the sea of Aramaic speakers. Even with the attempts to maintain separation described in 3.2.2, we would not expect more than a few generations of maintenance of Arabic in these cities.

3.5.4 Natural Disasters: Earthquake and Plague

To explain the contradiction between the ineffectual settlement policies of the Arab settlers, but the apparent success of Arabic, I argue that the primary engine of Arabization was actually a series of natural disasters which devastated urban and even rural areas in the Levant. This devastation created a vacuum, which allowed, or even forced, nomadic pastoralists to settle in the now vacant areas.

The Levant was hit extremely hard by plagues over a two hundred year period, beginning in the late Byzantine era (the 542 CE Plague of Justinian), and continuing to nearly the end of the Umayyad dynasty.²⁷ Though plagues also struck Egypt and Iraq, the Levant had a reputation for being beset by plagues — “the plagues of Syria” became a frequent topic in learned treatises, and numerous folk traditions evolved to explain why the Levant was so susceptible to plague — one folk tradition, for example, held that the prophet Muhammad banished plagues from Arabia, forcing them into Syria. A line of poetry states that:

Whoever reposes in Syria and spend there the night

²⁷ Conrad (1981, pp. 305-306) even suggests that plague might have been one of the factors that weakened the Umayyad dynasty and led to their defeat at the hands of the Abbasid revolution

Syria is grieved if it destroys him not. (Conrad, 1981, p. 331)

A chronology of plague epidemics in the Levant is shown in Table 3.1, along with the primary affected areas. This list is not of course exhaustive, but shows primarily those plagues which were recorded in literary sources and which are datable.

Year	Area influenced
542	Plague of Justinian: Antioch, Eiphaneia, Edessa, Jerusalem, Gaza, Ascalon; extensive in countryside
558	Antioch and elsewhere in Syria
573-74	Antioch
592	Antioch
17-18/638-39	Plague of 'Amwas: throughout Syria, especially Jordan and Palestine
53/673	Palestine
79-80/698-99	Syria: very severe, no raids against Byzantine Asia Minor this year
86-87/705-706	Damascus, elsewhere in Syrian countryside
94/713	Syria
99-100/718	Syria; central and northern regions
106-107/725	Syria; many villages destroyed
111/729	Syria
115/733-34	Syria
116/734-35	Syria
126/744	Syria; very severe, also famine
127/745	Syria; very severe in Busra and Hawran region, also famine
131/749	Syria; throughout area, also famine

Table 3.1: Plagues in the Levant. A reproduction of Conrad's (1981) Table 1 (p. 329): The Pandemic in Syria

The plague epidemics came to a end with the epidemic in 131/749, but this is the same year the Jordan river basin and surrounding areas was hit by a devastating earthquake. Monumental buildings were destroyed — two walls of the Aqsa

mosque collapsed and other churches and synagogues were flattened (Tsafrir and Foerster, 1992). Entire cities may have been destroyed, and ruins are clear in the archaeological record (MacAdam, 1994, p. 76; Tsafrir and Foerster, 1994, p. 111).

These types of natural disasters — plagues and earthquakes — affect cities and the countryside quite differently. For earthquakes, the difference is obvious — while an earthquake may be unsettling to someone living in a tent, the danger is minimal in comparison with the heavily developed cities of the late antique period, where buildings were built side-by-side along increasingly narrow streets.

Bubonic plague also strikes cities harder than the countryside. Since population density is one of the major determiners of the speed of the spread of a disease, epidemics affect cities more strongly than sparsely populated areas, with the vast majority of deaths occurring in spaces with closely placed dwellings and frequent interpersonal interaction (spreading pneumonic plague) and high rodent populations (carriers of flea-borne plague). Furthermore, plague tends to move from city to city, with the transport of goods and the plague bearing rats that move with them (Conrad, 1981, pp. 2-16).

Rural populations are also hit hard by epidemics of plague, but only after they begin in the cities. Epidemics seem to peak in March and September, key times for planting and harvesting, and remaining high throughout the summer (*ibid.*, p. 325). Depopulated rural areas would face significant losses of manpower, leaving fields untended and unharvested. A field left fallow for more than a year would begin to revert back to nature, and the investment in reclaiming it would be great.

Nomads, on the other hand, were relatively unaffected by these plagues directly, as the low population density of nomadic lifestyles, as well as their physical distance from centers of disease, kept them relatively safe. However, indi-

rectly, the devastation of the cities and the countryside indirectly affected their livelihoods, and creates a sort of ratcheting effect: As the disasters moved down the urban hierarchy, decimating first the most populous centers, then moving out to rural areas, the sources of agricultural goods which form the cornerstone of nomadic diets are no longer being produced by the former trading partners of the pastoralists. They are then forced to move into the vacuum and adopt a semi-sedentary lifestyle, a decision probably made easier with by the now fallow fields which could be used for pastureland.

Finkelstein and Perevolotsky (1990, p. 71) describe this breakdown of the symbiotic relationship between nomads and settled peoples very clearly :

The entire structure [of the symbiotic relationship] depends on the ability of the sedentary people to produce grain surplus for the pastoralists, since the sedentary groups are able to maintain a well-balanced nonspecialized economy, while generally the pastoral nomads cannot produce all the necessary basic commodities. A decrease in the agricultural production of the settled people will undermine those symbiotic relationships (no grain surplus for the nomads), and will cripple the desert dwellers' independence to practice "pure" pastoral nomadism. As a result, the latter are forced to shift to an agropastoral subsistence, in which each family supplies its own needs in both grain and animal products but does not produce enough surplus for barter. Such a process brings the pastoralists closer to full sedentarization, generally in the margins of the agricultural zones.

These disasters would therefore have created conditions which encouraged, or even forced, the Arabic speaking nomads to move into the areas which had

previous been the domain of sedentary Aramaic speakers. Following this, these increasingly settled Arabic speakers would have begun to move up the urban hierarchy, seeking wage labor in cities and eventually becoming urban dwellers, especially when cities were already heavily depopulated.

Another factor that promoted Arabization through these migrations was the new political situation following the conquests that created a military border between Anatolia and Syria. The new border would have biased migration in the direction of movements from the Arabian Peninsula into the Levant, as opposed to pre-Islamic patterns, where repopulation might proceed from the large Byzantine cities of the north, or even agricultural lands from elsewhere in the empire.²⁸

In spite of these calamities, the Levant was repopulated. Archaeological records show restoration of the damaged cities (Tsafrir and Foerster, 1994, p. 111), and there is a general resurgence in the fortunes of the Levant in the Abbasid period. Indeed, a model of population replacement (as opposed to just language shift) in the fertile areas of the Levant from desert dwellers is supported by the genetic data in El-Sibai et al. (2009) which finds a contrast between coastal populations, inland Levantine populations, and desert populations (probably the result of later population replacement in those low-density areas.)

Linguistically, it is not clear what effect the plagues would have had on those living in cities. Though the mortality rates were quite high, sometimes as much as 35%, the core of Aramaic speakers living in the cities probably would not have completely disappeared. It is difficult to imagine what effects they would

²⁸ Conrad (1981, pp. 485-6) notes that under Byzantine rule, plague-stricken areas would be repopulated from elsewhere in the empire. When Thrace was depopulated by plague, Armenians and Syrians were brought in to repopulate the area, and Constantinople was repopulated at one point by settlers from Greece.

have had on the Arab populations that had moved into these cities either — it seems likely that they would have gone from a small minority to an insignificant one, and certainly a minority whose linguistic impact would have been minimal. The presence then of proportionately larger Aramaic populations and still small Arabic populations is not a recipe for the linguistic success of Arabic, and so it seems likely that the cities of the Levant were Arabized in a slower process than in the countryside, and probably this Arabization significantly post-dated both the Islamic conquests and the Arabization of the countryside.

3.5.5 The Lebanese Mountains

While we have discussed above the settlement of coastal Syria and Lebanon, the difficult terrain of the mountains of what is now Lebanon (including both the Anti-Lebanon and Mount Lebanon) were probably slower to Arabize than the flatter terrain in Syria and Jordan. As with much of the Levant, this area was devastated by plague, earthquake, and then the Sassanian attacks in the early seventh century. Unlike the relatively easier terrain elsewhere, it seems like it would have been more difficult for nomads to penetrate into this mountain range. Instead, we appear to have more deliberate patterns of settlement.

The Ghassanids are said to have been among the first groups to settle Lebanon, with some modern Orthodox Christian families tracing their lines to the Ghassanids (Harris, 2012, p. 32). Similarly, the ʿĀmila tribe, originally said to have been living near the Dead Sea, moved into the mountains south and east of the Litani River, what is now called Jabal Amil in pre- or early Islamic times, and by the 650s CE were reported to be part of what was to become the Shiite movement. These settlers may well be the reason for the large Shiite presence in southern Lebanon.

There were still many Byzantine supporters in Mt. Lebanon and environs after the conquests, with the Byzantines sponsored Mardaites to maintain a Byzantine foothold. However, as the Islamic conquests consolidated their rule of the Levant, the Byzantines evacuated many of the Mardaites, and by the beginning of the eighth century CE, inland Lebanon was largely under Arab control (Harris, 2012, p. 37).

After the Abbasid revolution transferred the caliphate to Baghdad in the mid-eighth century, they felt a need to settle loyal, Arab Sunni Muslims to help control Lebanon. In 758 CE, the Caliph commissioned a branch of the Tanuḥ federation that at the time was living in northern Syria to move into the hills around Beirut, as well as around Mount Hermon, what is now on the SE border of Lebanon, Syria and the disputed Golan Heights. Their ability to colonize Lebanon was made easier by Christian revolts in the following two years, which were crushed harshly, and the rebels who had lived along the Beirut–Damascus were exiled, opening more space for the Tanuḥ. These Tanuḥ were able to create an independent empire in the late tenth century, and in the early eleventh they largely converted to become Druze, suggesting that this group had retained a community identity that allowed them to convert together (*ibid.*, pp. 39-40).

For an overview of the groups that were present in Lebanon, see Figure 3.4.

3.5.6 Summary: Arabization of the Levant

In summary, the linguistic history of the Levant is as follows: Aramaic was the primary language of settled communities in the arable heartlands of the region. Arabic was spoken by nomadic and semi-nomadic groups living on the margins of the agricultural viable areas. Greek was a language of government, inscriptions and learning, but does not appear to have penetrated very deeply as a language of

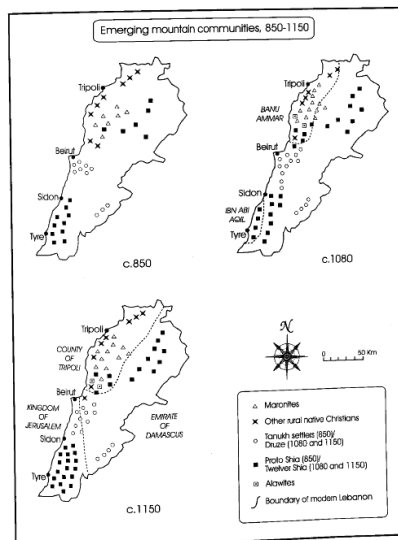


Figure 1.2 Emerging mountain communities, 850–1150.

Figure 3.4: Communities in Lebanon 850–1150 from Harris (*ibid.*, p. 41).

daily life.

The Arabization of the region can be dated quite clearly. In the early first century CE, we already see evidence of Arabic use in the area of the southern Levant and northern Arabia, suggesting that Arabic was in use as a spoken language even before this time. By the fourth century, the Hawran is clearly Arabized, though in most places Arabic is still restricted to the steppe. The Plague of Justinian in 542 CE is a likely starting point for the Arabization of settled areas of the Levant, beginning a two hundred year period of depopulation that finally ended in 131/749. These plagues and disasters would have created a drew Arabic-speaking peoples from the steppes into the fertile areas of the Levant. The dialects of these newly Arabized areas would have been local, coming from the Hawran and the nearby deserts, in contrast to those dialects that moved in later from the Arabian peninsula or possibly from Islamic Iraq.

The Arab conquests of the Levant would have primarily acted to wipe out the Greek speaking presence, as Greek speakers and those who had strong affiliations with the Byzantines, fled from the Islamic conquerers and moved into areas still held by the Byzantines. These conquests probably had relatively little direct linguistic impact. Their policy of settling directly in conquered cities probably led to rather rapid linguistic assimilation to Aramaic, and so the Arabization of Levantine cities was probably a much later process.

The one area where the Arabs did establish important new settlements in the Levant was along the coasts, where military garrisons were necessary to counter the Byzantine incursions by sea. We do not have a great deal of information about these settlements, but they might represent a somewhat separate layer of settlement from either the steppe-to-countryside movement, or whatever later process led to the Arabization of the urban centers. Similarly, the difficult terrain of the Lebanon mountains might have been colonized first in the south by proto-Shiites from the Jordan valley, and then in the areas along the corridor between Damascus and Beirut by Tanukh tribesmen from northern Syria.

3.6 Mesopotamia

The pre-Islamic social and linguistic situation in Mesopotamia and surrounding environs²⁹ is very similar to the situation in the Levant: at the edge of the region, in the deserts along the western banks of the Euphrates, a largely sedentary Arabic culture developed, in communication both with Arab desert pas-

²⁹ Defined as the region stretching from the desert fringe west of the Euphrates to the foothills of the Zagros mountains, thus including Khuzistan which was also Arabized, though it is now a part of modern Iran (now being Persianized, see Matras and Shabibi, 2007), as opposed to the highlands which likely remained largely Persian speaking.

toralists and the Persian ruled (but largely Aramaic speaking) people to the east of the Euphrates. As in the Levant, these Arabs were also recruited as a buffer state between the rival Byzantine and Persian powers, and were used to control and mobilize the desert tribes of the interior. The Arabic client state of the Persians, ruled by the Lakhmid's (sometimes called Nasrids) also lost its client status shortly before the Arab invasions, though these groups still resisted the newcomers.

In contrast to the Levant, the pattern of post-conquest settlement in Mesopotamia involved the foundation of entirely new cities, populated almost entirely with Arabic speakers. These settlements were likely the seeds of Arabization which were to inform the language of at least the urban centers of Mesopotamia.³⁰ Mesopotamia was generally less hard hit by plagues than Syria, and thus the newly established cities appear to have continued to act as centers of Arabic per the First Effective Settlement principle.

3.6.1 Pre-Islamic Linguistic Situation

Three languages predominated in pre-Islamic Mesopotamia. The language of the majority of the population living between the two rivers was Aramaic (in several mutually intelligible dialects, Beyer, 1986), while the ruling Persian elite spoke variants of Middle Persian and Arabs west of the Euphrates likely spoke Arabic. Aramaic was spoken widely, primarily in the fertile agricultural areas between the two rivers. Aramaic speakers also lived in cities, and even Madā'in, the conurbation making up the Persian capital, likely had a larger number of Aramaic speakers. In the north, Aramaic speaking agricultural areas gave way to Persian

³⁰ As per the Bedouin paradox, the language of the surrounding countryside is likely to be a more recent arrival, brought by Bedouin groups moving in from the desert. See for example (Palva, 2009).

and Kurdish areas, though some slave transfers may also have created an Aramaic presence east of the Tigris in those areas (Morony, 1994, pp. 170-1).

Middle Persian was spoken by the ruling elites, who resided largely in major urban areas, but also in garrisons and even in primarily Aramaic and Arabic speaking cities. Persians were differentiated from their neighbors in a variety of ways, which may have helped them preserve a distinct linguistic and cultural identity: they were an economic and administrative elite, and wore special clothing which reflected a complex hierarchy of status; they were largely Magian, as opposed to the Aramaic and Arabic speaking Christians and Jews. Multilingualism in Persian was not necessarily wide-spread, but it was a language of upper mobility and so Arameans and others who wished to join the ranks of the bureaucracy and aristocracy tended to become Persianized (Morony, 1984, p. 171).³¹

In the pre-Islamic era, Arabic was spoken primarily in the areas west of the Euphrates, though there were Arabic speakers who pastured their animals and even settled in villages east of the upper Tigris in the Sassanian Province of Beth ‘Arbhayē. As one moved further north, there was greater penetration across the Euphrates. It is not clear, however, whether these groups maintained the Arabic language — the rulers of Hatra, about 300km north of modern Baghdad have Arabic names at first, but we also find Persian names and Aramaic titles (even an Aramaic title translating to *malkā dī ‘Arab* ‘King of the Arabs’) so a certain amount of assimilation must have occurred over time (Bosworth, 1983, pp. 595-6).

As in the Levantine Hawran, there was a settled capital of an Arab kingdom, al-Ḥira, but there were many more cities along the Euphrates that seem to have

³¹ Persian appears to have been relatively successful as a language, and these attempts to differentiate Persians and their subject populations are close to the similar attempts by Arabs detailed above. However, as we have noted, the plagues probably reduced the viability of this strategy in the Levant, even if it might have been effective elsewhere.

been largely Arab, including Haditha, Hit and Anbar, as well as an Arab suburb of the capital of Ctesiphon. These Arabs were likely bi- or tri-lingual in Arabic, Aramaic and Persian, and were heavily influenced culturally by Sassanian society. For example, the system of government in al-Ḥira was clearly modeled on the government in Ctesiphon (Morony, 1984, pp. 221-2).

al-Ḥira was something of a predecessor of the later Islamic cities that would be established between the rivers in Islamic times. The city was founded around the third or fourth century by the Quḍā'a branch of the Tanūḥ federation, originally hailing from Yemen, via Bahrain (i.e. the NE Arabian coast). These tribes had begun penetrating past the Euphrates in the northern Jazira region, but were expelled by the new and powerful Sassanian state and some migrated to, and founded al-Ḥira.

al-Ḥira was less a cohesive city than a scattered, small collection of fortified palaces scattered among agricultural areas, and with a population probably no greater than about thirty thousand in the early seventh century (Kennedy, 2007, p. 104). Splinter groups of various tribes were settled in the city, generally keeping to their own enclaves (Donner, 1981, p. 183), though there was also an incipient shared ethnic identity between the Nestorian Christians in Hira, who were referred to not by a tribal affiliation but as ʿTbād (Morony, 1984, p. 222).

The importance of al-Ḥira was primarily in its role as the capital of the Lakhmid dynasty, which had operated as a client state of the Sassanians since the third century. However, shortly before the Arab conquests in 602 CE, the Sassanians broke off ties with the Lakhmids, reducing their prominence and forcing the nearby tribes to reevaluate their allegiances (Kennedy, 2007, p. 172). Thus, though the Lakhmid's may have formed a community that may also have been a speech community, it began dissolving roughly thirty years prior to the Islamic conquests.

Members of the Tanūḥ federation which helped found al-Ḥira were largely sedentary or semi-sedentary, while other Arabic groups lived on the fringes of the deserts. Many of these more nomadic groups seem appear to have lived relatively in-situ in the NE Arabian peninsula, including the Ṭayyi' tribal confederation, whose name was synonymous with nomads in the Aramaic of Iraq (*Tayyēyā*). These groups tended to be further south, closer to the Gulf and modern Kuwait, with the originally Yemeni Tanūḥ further north and more settled (Donner, 1981, pp. 170-171; Morony, 1984, pp. 217-218). tay

The kinds of interactions between the Arabs and other groups are shown by episodes related from the Babylonian Talmud. There are stories (ca. fourth century CE) in the Talmud of Arab thieves (obviously settled people living in an urban environment) as well as nomadic raiders, both told in relation to the town of Nehardea, well north of Tikrit on the Tigris, far from the deserts and the Euphrates, while in Pumbeditha, near modern Falluja on the Euphrates, Arabs were said to have forced people to turn over the deeds to their lands (an interesting choice of demand if this was a stereotypical 'nomadic' group). More peaceful interactions are recorded as well, with Arabs acting as market inspectors, and giving gifts to Jewish rabbis (Morony, 1984, p. 216). Again, we must be cautious about assuming that this ethnonym also referred to linguistic behavior.

The pre-Islamic linguistic situation can therefore summarized as follows: In the southern lands between the two rivers, Aramaic was the primary language, though Arabic speakers (likely originally from the southwest Arabian Peninsula) were present in the "Jazira" area between the two rivers in the north, mixing with the sedentary Arameans. Arabic speakers also were settled in towns along the banks of the Euphrates, where, though tribal ties remained strong, some were shifting to identifying themselves as a religious group. Further south and west of

the Euphrates, the Arabic tribes appear to have hailed from the eastern Arabian Peninsula. Cutting across these areas geographically, Persian speakers tended to occupy roles of prestige in both the city and the countryside, as government officials and landowners, and anyone aspiring towards service in the government would need to acquire Persian language and culture as well. However, all of these peoples were probably in frequent interaction, so that multi-lingualism would have been wide-spread, especially in areas close to the Euphrates, though it is likely that Arabs, as a politically and socially marginal group were multilingual in Aramaic and possibly Persian, while Aramean peasants were unlikely to be highly proficient in Arabic.

3.6.2 Conquest and Continuity

The Arab conquests in Mesopotamia were extremely rapid, and like the conquests in the Levant, quickly swept away the previous rulers. The Byzantine-Sassanian wars, which had resulted in the Sassanians occupying the Levant beginning in 611, swung back in the favor of the Byzantines who were able to penetrate to the doors of the Sassanian capital in 628, which triggered a succession dispute lasting until 632. A massive series of floods starting in 628 had disrupted agricultural life in southern Iraq, further impacting revenue and generally depopulating major areas that weren't repopulated until well after the Islamic conquests (Kennedy, 2007, pp. 101-103; Morony, 1984, pp. 190-1; Donner, 1981, pp. 172-3).

The Sassanians has in 602 CE ended their relationship with the Lakhmid Arab client-buffer state along the desert margins, and while this may have made the conquests easier for the Muslims, the Arabs living at the edges of Iraq were split in their allegiances when the Muslims came in 633. Some Arab tribes had already begun taking advantage of the power-vacuum and began asserting their

tribute rights to communities on the frontier, so that with the arrival of the Muslims they were able to parlay these activities into more active conquest (Kennedy, 2007, pp. 103-104). For the tribes that rebelled, they were treated as any other opponent, killed in battle with the survivors taken captive and assimilated into the structures of the conquering tribes (Morony, 1984, p. 223; Donner, 1981, p. 182).

However, many other Arab tribes decided to join the newcomers (sometimes with some persuasion), either by paying tribute as in al-Hira and other non-Muslim Arab cities or actively joining with the Muslim armies in search of booty and land. At times, members of the same tribe took different approaches to the invaders, and the lines which divided supporters of the Muslims and their opponents do not necessarily correspond to lineage or religious divisions within the tribe, underscoring the complex patterns of allegiance that make it hard to consider tribes as linguistic or even social units (Donner, 1981, p. 182).

In general, however, the Muslim conquests did little to change the lives of pastoralists at the edges of Mesopotamia. The conquests were not a general movement of pastoralists, but a concerted military campaign, and most of those participating in the conquest were settled in the new cities founded in Iraq, not in the already standing Arab towns. Thus, those Arabs living in cities and towns were largely unaffected, and while some members of tribes which rebelled against the Arabs were killed or captured, the bulk of the tribes remained essentially in their same range, especially in the south (Morony, 1984, p. 508). On the other hand, the conquests did make possible more movement from the Peninsula into the Syrio-Iraqi desert, creating a chain reaction that likely increased the Arabization of the areas north and east of the Tigris, in the areas around modern Kirkuk and Irbil, beginning in the second half of the seventh century (*ibid.*, pp. 230-231).

The primary group affected by the conquests was the Persians. Like the

Greek Byzantines in the Levant, the Persians had operated largely as a ruling class, occupying the upper reaches of a society which did not necessarily share their language or culture. The Arab conquests either killed off, captured, or put to flight, a significant portion of the ruling elite, including the royal family (Morony, 1984, p. 191). As in the Levant, once the tide of battle had turned (following the decisive battle of Qādisiyya in approximately 636) the ruling outsiders fled to their heartlands in Iran, leaving Ctesiphon abandoned and opening the way for a quick and easy conquest of the country (Kennedy, 2007, pp. 120-121).

Along with the ruling Persians, the conquests also swept away the state Zoroastrian/Magian religion. Prior to the Islamic conquests, Magians had already begun converting Christianity in larger numbers, including important members of the royal family and the aristocracy. However, the Islamic conquests seem to have dealt the coup-de-grace to the struggling religion. The property of the fire-temples was confiscated, since they were seen as part of the Sassanian government, whereas the property of other religious groups such as Jewish temples and Christian churches was largely left intact (Morony, 1984, pp. 298-300).

However, though large numbers of Persians left Iraq, some Persians found ways to stay and adapt to their new rulers, while Persian culture itself remained extremely robust. There were major military units of Persians who defected to the Muslim side and who helped settle the new cities in Iraq (Kennedy, 2007, pp. 131-132). Many of the Persian landowners also seemed to have cooperated with the Muslims in general. Continuing the trend of conversion to Christianity, many aristocratic Persians took important positions in the Nestorian Christian church or converted to Islam in order to keep their lands. However, by the eighth century, after prohibitions on Arabs settling outside of the *'amṣār* were lifted and the tax code began to change, the Persian landowners were being replaced by Mus-

lim Arabs (Morony, 1984, pp. 206-8). In general, as a social group, the Persians in Iraq began to be absorbed into Islamic Arab society by the beginning of the eighth century, with even the originally autonomous Persian military units being disbanded. This is not to say that Persian language disappeared, as there were still many Persian speakers insulated by the Zagros mountains against linguistic assimilation.

With regards to the Aramean population, the conquest essentially left them untouched, except for slight disruptions caused by the change in administration and an increase in taxation under the Muslims. This huge population was probably the numerically dominant demographic and linguistic group. In the area around Kufa, for example, the Aramean population is estimated at 1.5 million, while the Arab population was counted in the tens of thousands. It is likely that those Arabs who settled in the agricultural lands outside the main cities were quickly Aramaicized, as they had been pre-Islamically (*ibid.*, pp. 173-177).

Though the Islamic conquests essentially swept away the Magian faith, the new rulers maintained and extended a system that allowed religious groups a great deal of legal and social autonomy. Beginning in the fourth century under the Sassanians the Monophysite Christian, Nestorian Christian and Jewish communities in Iraq began to develop a system whereby the members of a religious community would enforce their own civil laws for issues such as marriage, divorce, inheritance, etc. Though the Arabs had not yet developed their own system for dealing with these communities, the individual communities acted towards their new rulers just as they had with the previous ones, and so the system continued. The Arabs therefore allowed these religious communities to continue to control their own affairs, provided they paid their taxes. They were very even handed in their administration, to the point that some religious figures actually complained that

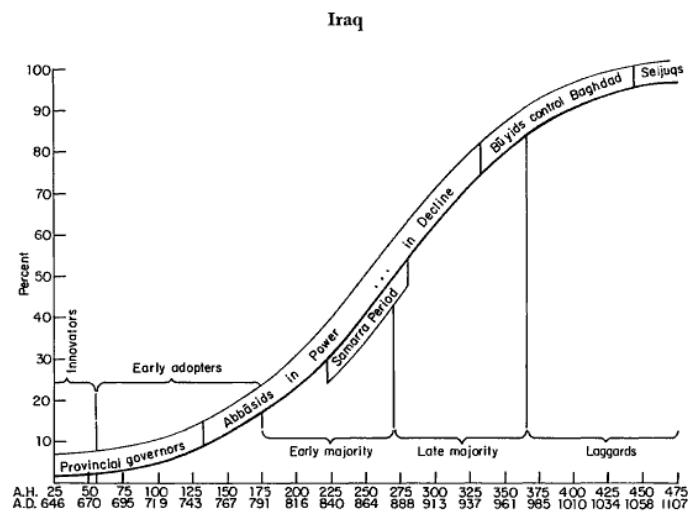
the Arabs didn't help wipe out their enemies (Morony, 1984, p. 346).

This continuity helped further a system that created strong community boundaries along religious lines, one which had begun three hundred years before the Arab conquests. Religious affiliation became an increasingly important marker of identity, superseding ethnicity since religious groupings and ethnic groupings did not necessarily coincide. Religious groups were already becoming increasingly segregated pre-Islamically, and the post-Islamic situation in Iraq probably did little to reverse this. A story from the life of the saint Abd-al-Masih reports that shepherds would not socialize with those from the other sect, and the young saint (not yet converted) was driven away by the Christian shepherds when he tried to eat lunch with them (*ibid.*, p. 314). At the same time, it seems impossible for anyone in a complex society to avoid interacting with members of different religions, and so it seems that the laity of these sects probably still conducted business with one another. We know, for example, that Christians pre- and post-Islamically attended Jewish taverns, as did Muslims later on (*ibid.*, pp. 371-2).

What this means linguistically is that the religious communal divisions which we know correlated with linguistic differences in the 1960s, per Blanc (1964), were in place in the pre-Islamic era. This social-religious divisions likely shaped the trajectory of Arabization as well as the development of the dialects that emerged from that process. On the other hand, we know from onomastic data (Bulliet, 1979) that conversion to Islam largely was complete in Iraq by the late 10th century (see Figure 3.5) and that Persians converted much more quickly to Islam than Iraqis in general, whether due to their participation in the *mawla* system (Crone, *EI2*, “*Mawla*”) or due to the incentives of conversion in retaining their land or in obtaining government positions (see Figure 3.6) (Bulliet, 1979, pp. 83-84). This suggests that within three centuries of the conquests, Christians and Jews were probably

a minority, and conversion to Islam would have subsumed members of various ethnic groups, probably changing the ethnic divisions into religious ones.

The other major linguistic outcome of the Islamic conquest of Iraq and the Levant was that it put the entire Greater Arabian Peninsula under an essentially unified rule. The old frontiers between the Sassanian and Byzantine empires vanished. Instead of a line bisecting the Jazira in Mesopotamia and down through the desert, the barrier was instead the contested Taurus Range between Byzantium and the Arabs. Movements between Iraq and the Levant would now be easier and freer. Indeed, movement from and to points further east would also be possible, since the conquests also brought modern Iran and parts of Afghanistan under Arab rule.



Graph 15. Correlation of historical developments with Iraqi conversion curve.

Figure 3.5: Graph of conversion to Islam against time in Iraq based on first-name onomastic data from (*ibid.*, p. 83).

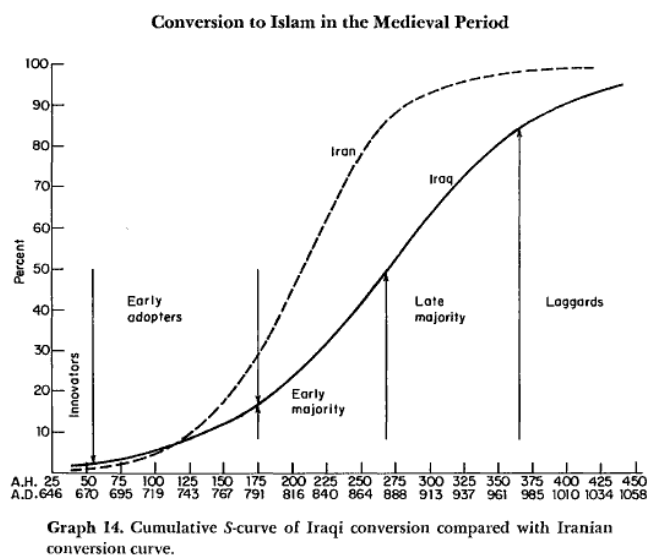


Figure 3.6: Graph of conversion to Islam against time in Iraq among local Iraqis and among Persians (Iranians) based on first-name onomastic data from (*ibid.*, p. 82).

3.6.3 New-cities founded in Mesopotamia

3.6.3.1 Kufa

Kufa was one of the first major established Islamic cities, and was to be one of the largest cities in Iraq until the foundation of Baghdad. Kufa was founded in ca. 638 CE/15AH by approximately thirty thousand veterans of the decisive battle of al-Qādisiyya.³² At most, six thousand of these were from the original veterans of the battles near Medina and Mecca, while nearly twelve thousand were from Yaman. The primary other residents (ca. eight thousand) were from North Arabian tribes (Nizārī) (Morony, 1984, p. 239; Donner, 1981, p. 229). Though the numbers are likely not entirely accurate, we get an impression that the majority of the original founders of Kufa were from the west coast of the Arabian peninsula.

The city was laid out, like most *'amṣār*, with a congregational mosque and government buildings at the center, and with quarters allotted to specific tribes (for a detailed breakdown of which tribes settled where, see Donner, 1981, pp. 228-229). One reason for creating special quarters for each tribe was administrative convenience — the new system of paying fighting men a stipend was enacted on a tribal basis, so that the payments were distributed directly to tribal leaders, who would pay it to their men (For the reasons behind tribal division, see *ibid.*, p. 234). This may mark the first time that many of these tribes came directly into contact with each other, and as Morony (1984, p. 234) notes:

The nature of the settlement itself, close together in a garrison city where tribal labels were used as a basis for organization, served to

³² Though figures vary, so that Donner (1981, p. 229) suggests a total military aged male population of closer to twenty thousand, and he notes that some sources suggest as many as one hundred thousand.

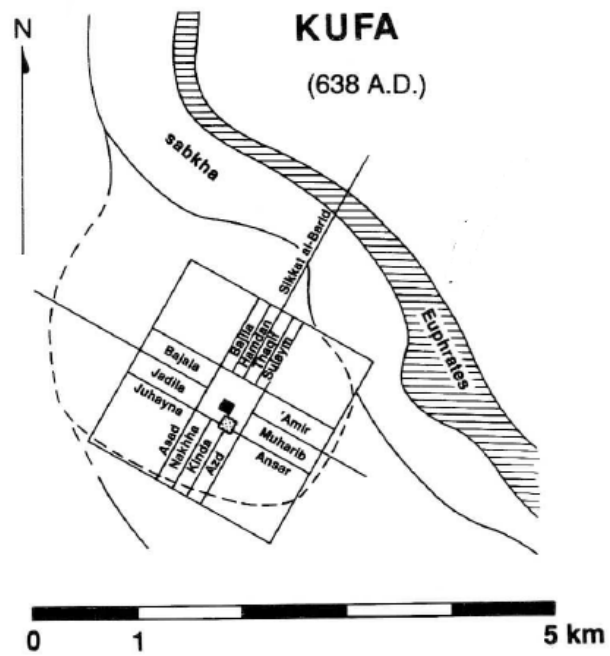


Figure 3.7: Reconstructed plan of Kufa, from (Whitcomb, 1994), figure 28. Black square is the central mosque, gray square is the administrative building.

intensify tribal identities, especially among groups that previously had only vague and distant, even fictive, ties of kinship. It also increased the opportunity for conflict along tribal lines. (see also Kennedy, 2007, p. 134)

Evidence of this increasingly tribal identification is the reorganization of the quarters when new members of a tribe would arrive. When new immigrants to Kufa came in groups that could not fit within the tribal quarter, the tribe would undertake a reorganization of the quarter — either by moving the entire tribe further out of the city, or by annexing lands left from by groups which had already done just that (Donner, 1981, p. 232). These quarters were in many ways self sufficient — many tribes established their own mosques, separate from the central mosque, which by default seems to have become part of the Quraysh/Hijazi dominated center of the city (Morony, 1984, pp. 239-240).

The growth of Kufa was driven initially by Arab immigration, hailing largely from the Hijaz and Yemen. Hijazi and Qurayshi immigrants seem to have formed something of an elite and were settled in the center of the city. And though Donner (1981, pp. 227-229) suggests that Kufa may have been established as a center to help settle pastoralists, many of these tribes seem to have been from settled areas in the Hijaz and agriculturally bountiful Yemen, so accounts which suggest this may have been part of a later ideology linking pastoralism to Islamic foundation myths.

There were also many Persians who moved into Kufa. There was a contingent of Persian defectors, *al-Ḥamrā*, who had fought in the Iraqi campaigns and who settled at Kufa, where they flourished, numbering nearly twenty thousand after a generation by about 680 CE, though some of that number was transferred

to live elsewhere (Morony, 1984, p. 198). Another significant group of non-Arabs were the captured Persians, as well as those Persians who had converted to Islam and became integrated into various tribal groups as (pl.) *mawāli*.³³ There were also many women who had been captured as their families tried to flee Iraq to Iran, and they and their children became an important part of Kufan life. Similarly, the movement of newly converted Muslims, especially Persians, (to be discussed below) likely also added some Persians to the city.

We can get some sense of the proportions of different groups in Kufa after one generation. In 657 CE, at which point Morony reckons growth was primarily caused by procreation rather than immigration, the authorities registered some 57,000 Arab fighting men, of whom some 40,000 were adults, and 17,000 were post-pubescent males capable of helping on the battlefield. Nearly 8,000 slaves and *mawāli* were also registered in the army (Morony, 1984, p. 244). If, as was the case in 645, there was at least one non-combatant for each combatant, this suggests a total population in Kufa in 657 CE of about 130,000 (*ibid.*, p. 243) or possibly as much as 200,000 (Djāit, 2007).

After the initial phase in which tribal affiliations were strengthened by moving into close proximity, the importance of the tribe as a indicator of social grouping slowly declined, as it had in al-Ḥira before. Though the distribution of wealth largely on tribal lines might have highlighted tribal divisions, we have indications that a general division between the wealthy elite and the bulk of the tribesmen began to emerge by the mid-seventh century. By the 680s, these elites had formed into a clear group, and were marked by their increased use of Persian

³³ A (sg.) *mawlā* was a free, freed or even captive non-Arab who was integrated into a tribe and treated as a member of that tribe, in much the same way that a person could get nationalized to a new nationality today (see Crone, *EI2*, “*Mawlā*”).

clothing and customs (Morony, 1984, pp. 256-260). We have reports of impoverish mobs attacking elites in 680 CE (*ibid.*, p. 262), and even in 656 CE the *Ḥawāriğ* rebellions seem to have been an expression of a “religiously justified opposition to the growing socioeconomic polarization in the garrison cities of Iraq (*ibid.*, p. 468).” Thus, a broader, more inclusive notion of social class seems to have been growing as an identity as early as the late seventh century, in contrast to the more fine-grained divisions of tribes.

There would have been limited exchange of peoples from other cities, such as al-Ḥira and the former Sassanian capital of al-Madā’in, both of which retained their importance for some time, and so there does not seem to be a scenario of migration to Kufa from these cities. al-Ḥira became something of a refuge for non-Muslim Arabs, though some people moved to Kufa without converting to Christianity, and al-Ḥira finally seems to have declined by the early tenth century (*ibid.*, p. 235). al-Madā’in also remained an important administrative center, and though it was sacked in 687 CE, this probably was restricted to Muslims, and the city remained viable until Baghdad absorbed it after its founding in the 762, which also led to the gradual decline of Kufa which was mostly complete by the end of the ninth century. Moreover, throughout the Islamic history of al-Madā’in, administrators and important Arabs were cycled between there and Kufa (*ibid.*, p. 251).

We therefore have the following possible linguistic scenario in Kufa. The city was largely settled by Arabic speakers hailing from the Hijaz and Yemen, though a significant portion of Persian defectors and captives also lived in the city. Tribal divisions would have been thrown into strict relief as a marker of identity in the first few generations, so that while members of a given tribe probably didn’t form a speech community prior to immigration to Kufa, they may have done so thereafter, as early as the second generation. In that second generation of

Kufans, the Persians who were captives would have had access to first-language acquisition of the dialects spoken by the households that they found themselves in, while the Persian from the al-Ḥamrā had their own division of the city and probably grew up as bilinguals given the environment. The primary population in Kufa would have been Muslim, while Christians and Jews continued to live in al-Ḥira, and Persians and Arameans lived in al-Madā'in.

Kufa was probably a major source of Arabization in the first century of Islamic Iraq, especially once the koineization process was complete. Even after the foundation of Baghdad 130 years after Kufa in 762, it probably continued to have a major linguistic impact as one of the first major Arabic speaking settlements in the country. Kufa didn't completely decline in importance until the beginning of the tenth century, at which point much of Iraq was likely Islamized (if not Arabized, see Figure 3.5).

3.6.3.2 Basra

Basra, like Kufa, was founded as a *miṣr* and may have been founded at the same time, or shortly after the founding of Kufa, ca. 15/635. In contrast to Kufa, it was founded by a significantly smaller population, numbering around one thousand fighting men, largely from the Hijaz, including the tribe of Quraysh.³⁴ However, a huge number of members of the tribal confederation Tamīm, which was likely from central and eastern Arabia, also settled in Basra, and by 656 CE accounted for 10-12,000 of the 60,000 fighting men registered the same year, and reports also indicate a number of other central and eastern Arabian tribes (Lecker, *EI2*, "*Tamīm b. Murr*"; Morony, 1984, pp. 246-250). The city grew rapidly, with a

³⁴ They are referred to as *ahl al-ʿāliya* 'people of the heights' and I've seen this refer to parts of Yemen. Need to remember where that reference is.

population of nearly two hundred thousand by 680CE, of which eighty thousand were men of fighting age registered with the army (Morony, 1984, p. 250).

Like Kufa, the city was divided by tribes, which also provided the divisions for military units. There was also a contingent of Persian cavalry, the *'asāwira*, originally an elite cavalry unit which occupied a similar place to the *al-Ḥamrā'* in Kufa. However, this group ended up on the wrong side of a civil war, and had their homes destroyed and some of their members deported in the 680s. Nonetheless, their identity seems to have remained intact much longer than this, as we have reports of a Persian-Arabic bilingual Quran scholar in ninth century Baghdad whose name suggests he was descended from the *'asāwira* (*ibid.*, 208, esp. fn. 165).

Basra, unlike Kufa, was not absorbed into the megalopolis that was Baghdad, and so likely the language of Basra and Basrans provided one of the major centers of Arabic linguistic gravity in Southern Iraq (and as the base of the Basran school of grammar, the center of standardization of Classical Arabic) for several centuries, until the city went into decline in the late twelfth century, ending up as a far-flung provincial city. The current city is some twenty kilometers distant from the original site, and owes much of its current size to growth in the twentieth century as it rose in prominence as a port and terminus for oil shipment, so it is not certain that there is linguistic continuity between the two settlements.

3.6.3.3 Baghdad

The city, or rather the scattered collection of settlements that constituted the city of Baghdad, was founded in 762 CE near al-Madā'in, the ancient capital cities of the Sassanian and previous Persian empires, as a new capital for the Abbasid empire. Though the primary founding population was Arab, a substantial portion of the population came from Islamized Persians and Arameans (Bulliet,

1979, pp. 86-87).

The foundation of Baghdad caused the three major cities in this area (from west to east, al-Ḥira, Kufa and al-Madā'in) to experience significant declines in population, which may have been due to migration to Baghdad. Speculatively, one wonders whether the Christian and Jewish populations of Baghdad are ultimately derived from the population of al-Ḥira (or they might be part of the al-Ḥira speech community) while the Arabic-speaking Muslim population may have been derived from Kufa, or at least their dialect would have been based on the dialect of Kufa to which their allegiances would be oriented.

Like other Arabo-Islamic cities, Baghdad was organized into quarters by ethnicity and occupation, and the accounts available emphasize tribal identities less than in Kufa and Basra, which may reflect a reduction of the importance of the tribe as an organizational category at this stage. The city was, however, quite scattered, with multiple, ethnically largely homogeneous and self-sufficient quarters. Baghdad grew extremely rapidly, and at the end of the tenth century it boasted an estimated one million inhabitants (Duri, *EI2*, "*Baghdād*").

3.6.4 Other settlements and settlement patterns in Mesopotamia

In terms of other urban centers, the city of Mosul seems to have been founded by an army based out of Kufa in 641 CE, and was also founded as a *miṣr*, though it was near a largely Nestorian Christian community (Kennedy, 2007, p. 137). Other (later) settlers at Mosul included members of the tribe of Azd, which seems to have also been from the Hijaz originally. There must have been other, minor towns and outposts settled by Arabs in Iraq, but we have little evidence for this in our literary and historical sources (Donner, 1981).

Even given the robustness of the first effective settlement principle, the

Arabization of Iraq is astounding. Kennedy (2007, pp. 137-138) estimates that in Kufa, even with immigration, the Arab population in the first generation did not exceed ten percent of the total population of the area. The system of founding *'amṣār* helped maintain Arab solidarity, and Arabs were further allowed to settle only on abandoned lands (Donner, 1981, pp. 243-244), preventing cultural and linguistic assimilation to Aramaic or Persian.³⁵ How then did such a small (even if rapidly growing) population cause a major shift in language, especially given the fact that later rulers of Iraq would be either Persian or Turkish?

The plagues which struck the Levant did not seem to play as major a role in Iraq as it did in the Levant. The plagues that struck Iraq were much less severe than in Syria, and while they did cause major loss of life, the cities in Iraq seemed to be growing quite rapidly nonetheless. In one of the last plagues that hit Basra, Conrad (1981, pp. 437-438) estimates a 25% loss of life, which still left the city with over one hundred thousand residents. The plagues might have disrupted agricultural laborers, as it did in Syria (*ibid.*, pp. 346, 473), but other factors probably played a greater role in disrupting country life than the plagues.

Even pre-Islamically, the Byzantine retaliation to the Sassanian conquests as floods and other catastrophes seem to have created a situation where many agricultural villages were uprooted and had to move elsewhere. Morony (1984, pp. 190-191), quoting Adams (1965), notes that late Sassanian and early Islamic villages are frequently found nearby to one another, but do not overlap, which “suggests that the Sassanian abandonment was associated with a social upheaval sufficient to break off the tradition of residence at most of the Sassanian sites (*ibid.*,

³⁵ When the Arabs did try to settle in occupied areas, it was not terribly successful, such that when the Caliph 'Umar tried to settle Arabs from Najran in an occupied village in the fertile Sawad, they were forced to move and find another place to live (Morony, 1984, p. 252).

p. 223).” However, nearby sites were settled with similar population density shortly thereafter, such that life in these areas was disrupted, not destroyed, at almost exactly the time that Arab speakers began to make significantly greater inroads into Iraq.

The Islamization of Iraq also significantly disrupted the lives of new converts to Islam. Muslim converts typically moved to new settlements where they would be accepted, whether major urban areas such as Kufa and Basra, or even small villages where Muslims were the majority (Bulliet, 1979, p. 87). Bulliet (*ibid.*, p. 87) even argues that major cities such Baghdad and (its later sister city) Samarra lost population once the mid-point of conversion to Islam was reached, as the countryside now provided just as good a haven for Muslims as did the cities.

Muslims were, as Morony (1984, pp. 431-2) notes, a society of converts, encouraged to break off ties with Christians and Jews, and legally separated from those communities by the legal system of communal division. New converts to Islam had their lives significantly disrupted, and as Morony (*ibid.*, p. 273) notes, “[a]ssimilation was most complete for individuals who were removed from their former social contexts and integrated as individuals into a new society.” Given the curve of assimilation in Figure 3.5, the need for new Muslims to move to an urban area was probably highest in the first three hundred years after conquest, a period that also coincides with the peak of Kufan and Basran vitality.

At the same time, the rural agricultural practices seem to have been reoriented from subsistence agriculture to producing commercially viable products for the urban markets (Morony, 1994, pp. 228-229), a change which would have increased the links between the countryside and the city, economically and socially, especially if farmers needed to travel to the urban areas to trade in their markets in the dominant language.

This all contributed to a general urbanization of southern Iraq, with attendant Arabization: The incredibly growth of these new urban centers, both from procreation and immigration, combined with the disruptive effects of the Arab conquests and conversion to a new religion, and a shift to an urban-oriented economy. As speakers of Aramaic moved to those cities, or found their lives disrupted by the changing agricultural patterns, they would have been under ever pressure to acquire the Arabic language, especially if they moved into a primarily Arabic-speaking urban environment.

3.6.5 Summary: Arabization of Iraq

The Arabization of Iraq, like that of the Levant, began on the desert fringes, with settled Arabic-speaking groups well established west of the Euphrates prior to the Islamic conquests, and in the northern Jazira area, some ethnic Arabs coming close to the Tigris. The Islamic conquests established new, ex-nihilo cities, the *'amṣār*, which helped the new Arab settlers resist the dangers of assimilation by the much larger groups of Aramaic speakers that surrounded them, while urbanization and the lifestyle disruptions caused by conversion to Islam likely helped realign the speech communities in Iraq to include most of the population in the Arabic-speaking community.

Kufa and Basra were both founded by largely Hijazi and Yemeni populations, the speech communities which may have contributed the majority of the linguistic material to the Arabic of Iraq. Basra however also had a significant population from central and eastern areas of the Arabian Peninsula, and so the dialect of Basra and the areas around it might allow us to have some view of features from that area of the peninsula in pre-Islamic times, though it was never a large city and may have undergone various linguistic changes. Baghdad was likely founded by

an Arabic-speaking population drawn from Kufa and al-Ḥira, the latter probably contributing the Christian and Jewish dialects, which might give us some clue about the dialects of the desert fringe prior to Islam.

In terms of social groupings, the settlement patterns in Iraq reified the divisions between tribes which may not have corresponded directly to linguistic divisions prior to those settlements. Whether linguistic differences broke down along tribal lines later on is unclear — the move towards social groupings that cut across tribes (such as social class and religion) was fairly rapid, and sources on the settlement of Baghdad emphasize tribal organization somewhat less than those about Kufa and Basra. The Sassanian system of dividing religious communities, who themselves tried to create exclusive, inward-facing groups, continued following the Arab conquests. The fault lines of religion may have determined the course of Arabization of those groups, which may be reflected in the language, and we know that linguistic divisions can ‘break’ along those lines, as with the Muslim dialects of major Iraqi cities which have become significantly Bedouinized vis-a-vis their non-Muslim neighbors (Blanc, 1964; Palva, 2009).

3.7 Summary: Arabic and the Greater Peninsula

Within the Greater Arabian Peninsula, we can see that Arabic was well established prior to the Islamic conquests. The Arabization of Yemen probably began quite early, at least by the turn of the era. Ancient North Arabian speakers appear to have been shifting to Arabic for a long time, and certainly by the fourth century CE that shift seems to have been largely complete. The range of Arabic in the northern half of the Greater Peninsula was probably still restricted primarily to the steppe even at the dawn of Islam, with the exception of the client kingdoms who were permitted to establish themselves in more fertile settled areas than most

nomads. The plagues in Syria, starting in the sixth century, probably allowed Arabic speakers to make inroads into settled areas (but outside of the Hawran) even before the coming of Islam.

The Islamic conquests themselves primarily resulted in the removal of the previous imperial languages of the region, Greek, Persian and OSA, but did little at first to disrupt the daily lives of the largely Aramaic speaking (in the north) and OSA and MSA speaking peoples in the south. The conquests established centers of Arabic speakers in Iraq, but probably did little to establish effective urban, Arabic-speaking settlements in the Levant. Instead, the Levantine countryside was probably opened up for Arabization by the plagues, meaning the countryside was Arabized before the cities, the opposite of the pattern in Iraq. The conquests also united the entire Greater Peninsula under Arab rule, which likely made movement between the Levant and Iraq easier, as it was no longer a frontier zone.

Chapter 4

History: Outside the Greater Peninsula

This chapter applies the same historical, socio-linguistic analysis to those areas immediately outside of the Greater Arabian Peninsula. Though there was possibly some Arabic-speaking presence in Egypt pre-Islamically, Arabic became the dominant language in these regions only after the coming of Islam.

4.1 Egypt

The Arabization of Egypt seems to have been undertaken largely by Yemeni tribes, with some from Syria as well, who established the city of Fustāṭ which was absorbed later into Cairo. The Arabization of Egypt may have been helped by depopulation wrought by plague and warfare immediately preceding the conquests, and later by the establishment of the Fatamid state which helped move it conceptually and politically from the periphery to the center.

Throughout its history, Egypt had largely shrugged off the linguistic influence of many of its invaders, likely due to its high population density (Ostler, 2005).

However, immediately prior to the conquests, the Egyptian population seems to have declined drastically. In Roman times, its population was probably nearly five million at its peak, but plague had reduced it to closer to three million in the later Roman period, the first plague striking in 541 CE. Moreover, in the early seventh century, the Persians pillaged Egypt as far south as Aswan, leaving the land “half-empty” (Kennedy, 2007, pp. 141-143). Even when the Byzantines reestablished their rule a decade later, their persecution of the heterodox native Coptic church kept the nation in something of a state of conflict.

Linguistically, most Egyptians were likely native speakers of Coptic, an Afro-asiatic language descended from ancient Egyptian, though Greek was used widely as a language of writing. The extent of speech in Greek has been debated, though the testimony of papyri suggest that Greek monolinguals did exist in pre-Islamic Egypt, and there were likely many bilinguals also (Papaconstantinou, 2012, p. 63; al-Sharkawi, 2010, p. 153). Nonetheless, Greek was likely marginal in the lives of the vast majority of Egyptians (al-Sharkawi, 2010, p. 158), and Latin and Romance knowledge would have been minimal.

There may have been some Arabic speaker present in Egypt from well before Islam, though likely those who settled in densely populated areas, like the Delta, were linguistically assimilated. Herodotus, writing in the fifth century BCE, writes of an Arab nation living between the eastern Nile delta, the Sinai peninsula, and southern Palestine, an area that would be later occupied by the Nabateans (Retsö, *EALL*, “*Arab*”). Unfortunately for us, these lightly populated areas are also subject to rapid replacement of population, so Egypt dialects were probably not heavily influenced by these earliest settlers.

4.1.1 Settlement of Egypt

The conquest of Egypt was rapid and reached the fullest extent it would for centuries by the end of the fourth decade of the Islamic era. The conquest began in the 18th or 19th year of the Islamic era (ca. 639CE) and the southernmost point of the conquest, Aswan, was reached by 31 AH/651-2 CE. As in Iran and Syria, the most immediate consequence was the removal of the pre-Islamic administration. When Alexandria was captured in 641 CE, for example, the terms of the agreement required that the Roman army return to Constantinople by sea (Kennedy, 2007, p. 159).

The conquering army consisted largely of Yemenis. Kennedy (*ibid.*, p. 147) tells us that the initial army was made up of some 4000 troops from the Tihama plain in western Yemen, from the tribe of ʿĀkk. Later reinforcements increased the total size of the armies to perhaps twenty thousand troops (al-Sharkawi, 2010, p. 161), but still the vast majority of these troops seem to have been of SW Arabian extraction (Kennedy, 1998a, p. 64)

The primary settlement in Egypt was the new-city *miṣr* of Fuṣṭāṭ, founded a short distance away from what is now modern Cairo (much like Kufa is a short distance from modern Baghdad), and which became a major administrative capital and launching point for military endeavors further west. Though Alexandria had long been the primary city of Egypt, its importance became much more marginal as Fuṣṭāṭ increased in power, though it did host an Arab garrison and was used as a base for naval operations in the Mediterranean.

Fuṣṭāṭ was founded in ca. 22AH/643CE near the old Roman city of Babylon, possibly on the site of the original siege encampment, and like Kufa was a somewhat loose collection of settlements, separated at first by large empty spaces, and loosely organized by tribe though not quite as strictly as in Kufa or Basra, with

much more frequent movement of groups between quarters. The central structures were as usual the primary mosque and the administrative building, with the elite settled nearest to those buildings. These elite were primarily those closest to the Qurayshi, Meccan establishment, primarily from the Hijaz, but may have included the initial group of Tihman Yemeni from the tribe of ʿAkk, since they are not mentioned separately, that is to say that even in the prestigious center of the city, southern Yemenis may have predominated.

The other Arab settlers were, like the conquering army, largely from the south-west peninsula. Kubiak (1987, p. 69) estimates that at least three quarters of the army was from southern Arabian groups, though this is a large area and may have included speakers from quite different speech communities. Furthermore, and contrary to assertions of al-Sharkawi (2010, 164-169ff.), non-Arabs formed an important part of the initial settlement, though their quarters were often located at the periphery. There were contingents of Byzantines, probably part of the original expeditionary force from the Levant, converted Jews of unclear origin, Persian auxiliaries from Yemen, as well as local Copts who may also have functioned as auxiliaries or as administrators capable of handling the Byzantine records (Kubiak, 1987, p. 63).

Later settlement would likely have included the families and dependents of the conquering soldiers, as well as soldiers interested in joining the westward conquests, or those being resettled by the authorities elsewhere (*ibid.*, p. 78). Within the second generation, by the rule of the caliph Muʿāwiya (41-60AH/661-680, i.e. approximately 20-40 years after founding) there were 40,000 fighting men registered with the state for dispensation of stipends, though it seems likely that this was capped and there could have been more (Kennedy, 1998a, p. 65; Kubiak, 1987, p. 82). The number also does not include dependents or slaves, so the total pop-

ulation must have been much higher. Kubiak (1987) estimates that by 750CE, i.e. some four generations after foundation, even with conservative reckoning, there would have been nearly two hundred thousand residents of Arab descent, without including growth from non-Arab immigration and slavery.

The increase in population would also have included native Copts, who would have worked as tradesmen, traders, etc, catering to a population whose income largely came from the state through the *diwān* system of maintaining the army and its dependents. Though there were, initially, restrictions against the settlement of non-Muslims in Fustāt, by around 47AH/667 CE (i.e. about 25 years after foundation) this seems to have been relaxed, and the Christian population (probably largely Coptic Christians moving into the spaces between individual quarters) seems to have increased significantly. There would also have been a large number of slaves, whose numbers were regularly added to the settlement, whether Berber speaking slaves from North Africa, or Nubian slaves (400 of which were sent from Nubia each year), and indeed the first market in Fustāt appears to have been a slave market (*ibid.*, p. 81).

It is unclear what influence, if any, this immigration would have on the linguistic behavior of those living in Fustāt. The strong Arab character of the city, combined with the scattered settlement patters of the new immigrants, or their integration directly into families, suggests that native acquisition of Arabic would be quite possible within a single generation.¹

Arab settlement outside of Fustāt is less clearly documented. There were some Arabic speaking and writing landowners as early as 735CE, though it is un-

¹ This contrasts with the thesis of al-Sharkawi (2010) that colloquial Egyptian Arabic is derived from the foreign-talk used by Arabs to speak to the quite separate Copts. However, he ignores most of the evidence of social integration within a dominant Arab context suggested by Kubiak (1987), which would favor native acquisition and shift.

clear whether they lived on their lands or were absentee landowners (Papaconstantinou, 2012, p. 68), and the pace of agricultural settlement was likely slow at first, with the usual restrictions on settling or purchasing occupied land. We know there was a garrison in Aswan with supposedly 20,000 troops around the third Islamic decade (al-Sharkawi, 2010), but given that a peace treaty was finalized with the Nubians the following decade, and that wars were ongoing in North Africa to the west, it seems likely the troops in Aswan may have been moved around later.

The primary populating group in Egypt outside Fustāṭ seems to have been from tribal groups not from Yemen. The first such group was some sort of tribal confederation known as the Qays, whose origin is unclear but may be from the area stretching from the Hijaz to the Western Najd. They probably participated in the conquest of Syria and settled there for a time.² The Qays has participated in the civil war against the Umayyads, taking the losing side of the Zubayris, participating in the defeat at Marḡ Rāhiṭ in 65/684 (where the victors were said to be Yemeni tribes). However, they may have switched sides in that battle (Elisséeff, *El2*, “*Mardj Rāhiṭ*”), and forty years later in 109 AH/727 CE, we hear that the Qaysis are beginning to be being settled in Egypt as a counterweight to the Yemeni tribal contingents in Fustāṭ. This suggests that in the generation or two between these events, the Qaysis were back in favor with the Umayyad rulers, and that they were seen as a more reliable supporters of the Umayyads than the restive Yemeni elements in Egypt.

A generation later, however, in the 130s/740s, the governor of Fustāṭ forcefully expelled all of the Syrians (which seems to have included the Qaysis) who had

² We know, for example, that the two tribes Banū Sulaym and Banū Hilāl (who are important later for the Arabization of North Africa) hailed from the area between Mecca and the Najd, and that they formed an important component of the initial settlement of Qaysis in Egypt (Sajīnī, 2007, pp. 28,34)

migrated into the capital, who then probably settled in either the Eastern Ḥawf or Upper Egypt. In any case, this seems to indicate that the Upper Egypt's earliest major settlers were from this Qaysi contingent, rather than the Yemeni conquerors. Indeed, Sajīnī (2007, p. 33) notes that a part of the contingent of Banū Sulaym in Egypt settled in Aswan, though the total number of Qaysis in Egypt according to his sources seem to be rather low, some 5,000 by 143/760 (Kennedy, 1998a, 75, arrives at the same figure).³ Nonetheless, these two groups, the Yemenis and the Qaysis seem to have had a political rivalry which defined the politics in Egypt in the first half of the second Islamic century.

This rivalry eroded fairly quickly with time. By about sixty years after the initial settlement of Qaysis, we have reports that the Qaysis and Yemenis were united by that most universal of common foes, paying taxes, and rebelled as a combined force against the governor in 168 AH/784 CE, and again in 214/830, and finally in 216/831 when they also joined forces with angry Copts.

For Alexandria, the garrison was stationed in a *ribāṭ*, some sort of fortified structure within or near the city, and was regularly replaced at six-month intervals, though some Arab soldiers were permitted to settle in houses abandoned by the fleeing Byzantines, similar to the situation in Damascus (Kubiak, 1987, pp. 76-77). Prior to the Islamic conquests, Alexandria had been the second largest city in the Byzantine empire, though following the conquests a large portion of the Greek population seems to have fled back to the empire. Alexandria, however, lost much of its importance, and may well have been rather secondary to Fustāṭ, though it did enjoy a certain amount of political autonomy until it was integrated into Egypt

³ Sajīnī (2007, pp. 32-33), contra Kennedy (1998a, pp. 74-5) refers to these settlers not as *al-qaysiyya* but rather as *ahl al-bayt*, i.e. descendants of the prophet and therefore presumably members associated closely with the Umayyad elite with a Hijazi extraction.

and bound more tightly to Fustāṭ under the Ṭulūnid dynasty (254–92 AH /868–905 CE), a policy that was continued under the Fatimids. The early use of Alexandria by the Arabs was primarily as a naval base, employing Coptic artisans who had experience with shipbuilding. It seems unlikely that the Arabization of the city began in earnest until well after the initial conquests, perhaps under the Fatimids, as discussed below.

4.1.2 Arabization of Egypt

The initial conquests placed maybe 100,000 Arabs in a sea of some three million Coptic or Greek speaking native inhabitants. However, by founding a major, and rapidly growing city that was separate from previous centers of power, the Arabs were able to resist being subsumed by the local population. Moreover, since most of the initial Arab inhabitants were financially independent due to payments from the central government to the military, any interactions they had with the natives were probably commercial, with the balance of power in favor of the Arabic-speaking conquerors. Fustāṭ would have been a large, strongly Arabic-speaking urban center, and even groups which were not Arabic speaking likely had the opportunity to acquire the language natively within a generation or two.

The relations between Arabs and Copts were initially ambivalent. The Arabs were fighting to remove the Byzantines, who had become increasingly oppressive to the Coptic church, but they were also conquerors. There are frequent reports in the literature of Copts aiding the Arabs during the conquests, and after the conquests we know Copts were employed in shipyards, building the Arab navy (Kennedy, 2007, pp. 352-354). At the same time, the taxation to support military campaigns and the expenses of caliphate ruled from Damascus and then Iraq led to serious Coptic rebellions beginning around 107/727, and continuing on and off

until the early ninth century. As discussed above, however, tax rebellions were not limited to a single social group.

The Arabicization of the administration of Egypt was ordered in 87/706, making Arabic the language of upward mobility (see Papaconstantinou, 2012, 67, for examples). As soon as a century after this time, starting in about 800 CE, legal documents, even for Christians, are almost entirely in Arabic, and those Coptic documents that have been preserved show a “laconic brevity and poverty in clauses (Richter, 2009, p. 421)”. Analysis of loanwords in Coptic documents shows that most loanwords are for luxury goods and legal terms, suggesting a great deal of economic intercourse, and that Arabic culture was the producer and distributor (as well as a consumer) of luxury goods (*ibid.*, p. 433).

On the other hand, the watershed moment for the Arabization of Egypt seems to have been the change of Egypt from a province of the Caliphate, centered in Baghdad, to an autonomous power in its own right. Drawing on Mufwene’s (2004) notion of settlement vs. exploitation colonies, Papaconstantinou (2012, pp. 73-75) suggests that the gradual autonomy of Egypt, primarily under the Fatimids beginning in the second half of the tenth century changed Egypt from an exploitation colony to a settlement colony, when the wealth of the country was no longer exported east to the Umayyad or Abbasid caliphates.⁴ This change made Egypt the center, rather than the periphery, and created a new, group identity centered around Egypt, but one associated with the use of Arabic:

The new situation not only enhanced the prestige and usefulness of Arabic, it also had some bearing on the very important question of

⁴ Though one might suggest that this process began even earlier with the Tulunids, starting ca. 250 AH/ 868 CE (Bulliet, 1979, p. 98).

group identity. If language is a vehicle through which a group can express its affiliation to an ideal community, or its indigenous roots as opposed to the foreignness of other groups, one cannot avoid the inference that this function was now filled by Arabic, and that this was partly the result of the Fatimid system. Here was a dynasty that could be identified as distinctly Egyptian thus quite plausibly filling the need for Egyptian self-identification, pride and symbolic capital. (Papaconstantinou, 2012, p. 75)

This autonomous turn in Egyptian identity also seems to have corresponded with the development of a majority Muslim population. Bulliet (1979, pp. 97-98) suggests that the first attempts at independence from the Caliph arose when about one third of the population had converted to Islam, corresponding with the short-lived Tulunid rebellion starting in 255AH/868CE, and ending with the reabsorption of Egypt back into the Caliphate in 292/905. This bid for independence may reflect the feeling on the part of the Muslim population that they were “sufficiently entrenched in Egypt to be willing to cut the umbilical cord uniting them to caliphal authority in Baghdad (*ibid.*, p. 98).” The final, successful movement towards autonomy via the Fatimids starting in 969CE happened at the point of two-thirds conversion to Islam, which was true of other autonomy movements elsewhere in the Islamic world.

This is also the same era that witnesses more evidence of wide-spread Arabization, at least in the populous and probably more thoroughly Arabized north. Though the Arabization of the administration was ordered in 87/706, the Coptic Church began its Arabization after moving to the Fatimid capital of Cairo in the early eleventh century, about fifty years after the foundation of that city in 974CE. It is also at this time, the early eleventh century, that we find an angry screed

against the use of Arabic, written by someone claiming to be a seventh century monk:

They are abandoning their beautiful Coptic language [...] and they are teaching their children from infancy to speak the language of the Arabs [...] even the priests and monks — they as well — dare to speak in Arabic [...] What shall I say in these times, when readers in the Church do not understand what they are reading, or what they are saying, because they have forgotten the language? These truly are miserable people, deserving of being wept over, because they have forgotten their language and speak the language of the *hiğra* [Arabic, ‘migration’]! (Richter, 2009, p. 426)

Fatimids rule also integrated Upper Egypt more closely into the trade networks of their empire. The most important route went through Qus to the Red Sea, which likely led to greater Arabization of Upper Egypt as traders moved through here (Papaconstantinou, 2012, p. 75). At around the same time as these trade routes were growing in importance, the Banū Hilāl’s rebellion was put down by the Fatimids, who exiled them into Upper Egypt (368/978) where they remained for approximately a century before they moved into North Africa (beginning 439/1047) (Schleife, *EI2*, “*Hilāl*”), and where the undoubtedly must have had some linguistic impact on the areas in which they settled.

4.1.3 Summary: Arabization of Egypt

It seems that the Arabization of Egypt began in Fuṣṭāṭ, later Cairo, where a mixed group of Arabic-speaking settlers, the majority of whom hailed from Yemen, founded a major city that would operate as the center of Arabization. After about a

century, a group of Arabic speakers from the Hijaz and possibly the Najd, perhaps by way of the Levant, were settled to the NE of Cairo in the Eastern Ḥawf, as well as in Upper Egypt, which extended as far south as Aswan. The process of language shift would have proceeded slowly until Egypt started to assert itself as an independence state, a process which began in mid-third/mid-ninth centuries, and which was completed with the coming of the Fatimids in 969CE.

4.2 North Africa

Unlike the conquest of other areas, the conquest of North Africa proceeded in a number of phases, few of which brought the kind of intensive settlement seen elsewhere, and as a consequence the Arabization of North Africa was slower and less complete than in other areas conquered by Arabic speakers. The native Tamazight (Berber) languages are still quite strong, and bilingualism in Arabic and these languages is not unusual from Morocco to the Siwa oasis in Egypt.⁵

Like many other areas that were conquered by the Arabo-Islamic armies, North Africa had been in something of a decline prior to the seventh century CE. Until the fifth century, North Africa had been tightly integrated into the pan-Mediterranean trade networks, sending olive oil, grain and “red slip” pottery north and east to markets under the control of the Roman empire. However, the Vandal conquest of Carthage in 439 CE seems to have significantly disrupted this trade, and even the Byzantine reconquest of 533 CE did not seem to reverse the trend. By the eighth century the archaeologically diagnostic red slip pottery seems to have completely ceased to be manufactured. Political events also weakened North Africa militarily. Heraclius, based in North Africa, rebelled against the Byzantine

⁵ For more see the *Berber Sociolinguistics* edition of the International Journal of the Sociology of Language, 1997, issue 123.

authorities in Constantinople in 610 CE and when he was successful, he seems not to have replaced the veteran troops he brought from North Africa to Constantinople (Kennedy, 2007, p. 203).

The economic decline seems to have significantly reduced the density of settlements in North Africa in the pre-Islamic era. As early as the third century, large parts of North Africa were outside of Roman control, and in many places cities seemed to be devolving into villages, partially due to being outside of the areas invested in by the central government, and also due to cities acting as Roman bases of power, which was now being actively opposed. By the seventh century, even Carthage, the capital city of Roman North Africa, was partially in ruin. The countryside was not spared either, and pastoralism seems to have increased at the expense of agriculture. Villages previously had been built on open plans with no walls, but at this time, to offer protection from raids, they began to be organized around fortified farmsteads not unlike those in the Arabian Peninsula (*ibid.*, pp. 202-4).

The primary population in North Africa were Berber (Amazight), speaking an Afro-Asiatic language, and pursuing a variety of life-styles, from pastoralists to sedentary agriculturalists to merchants and city dwellers. In many cases, they were organized in a tribal system, though our documentation is not particularly good and doesn't indicate whether this was true of all subsistence groups. The Arabic sources do seem to treat large Berber tribal confederations in an analogous manner to their treatment of Arabic tribes.

The other major population, especially in major cities, would likely have been Romance speaking, with some Greek speaking Byzantines also present. The area around Carthage, encompassing much of modern Tunisia, had been a colonized Roman province from 46 BCE until the Vandal conquests in 439, and it

seems likely that there would have been a significant number of native speakers of Romance, and some Greek speaking Byzantine soldiers immediately prior to the coming of Islam.⁶

There were roughly three phases to the Arab conquests of North Africa. The first was the campaign of ‘Amr b. al-‘Āṣ, the same leader who had conquered Egypt with his largely Yemeni army and who then moved on to North Africa beginning circa 642 CE. This set of military actions were closer to a raid than a war of conquest, and did little to establish a meaningful Arab presence in North Africa, with the exception of the Libyan city of Barqa, where ‘Amr imposed taxes and established a governor. However, ‘Amr was dismissed shortly thereafter for political reasons in 645, replaced by ‘Abdallah b. Abī as-Sarḥ, who was supplemented by a new army, some 5-10,000 strong, likely also South Arabian troops. The primary result of this campaign was the defeat of the Byzantine army in North Africa at Sbeitla, which retreated to Carthage and left their Berber allies to fend for themselves (Kennedy, 2007, pp. 206-8).

The second phase of the conquests established the first substantial presence in North Africa. It was led by a member of the following generation, beginning more than twenty years after the first, and led by ‘Amr’s son, ‘Uqba b. Nāfi‘. Beginning in 669, ‘Uqba led a mixed force of troops from Egypt, and Berber’s who had converted to Islam, with some 10,000 Egyptian troops (Kennedy, 2007, p. 209;

⁶ I wonder whether Carthage/Tunisia, which seems to have been the focus of Roman colonization, didn’t have a significantly larger amount of non-Berber speaking population, so that when the Arabs severed the ties with the Byzantines and the Romans, these would have been faster to switch languages, whereas in other parts of North Africa, there was more autonomy and less Roman settlement. This might explain the relatively stronger Arabic and weaker Berber presence in Tunisia. See for example Heath (2002, pp. 3-5) who argues that much of Morocco’s northern coast was likely Romance speaking, though some of his other conclusions about the process of language acquisition are unconvincing.

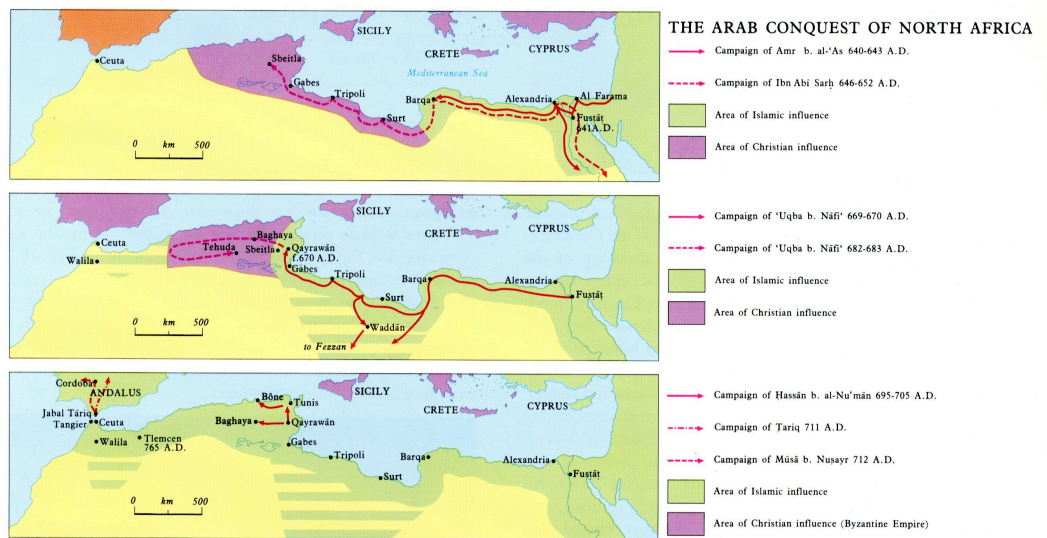


Figure 4.1: Conquests of North Africa from Ajaya and Crowder (1985). Note that the terms “Christian” and “Muslim” influence are misleading, and would be better termed Byzantine and Caliphate since the religious orientations in these areas changed slowly.

Ṭāhā, 1989, p. 61). He had good timing: the Byzantines were undergoing a dynastic struggle and the Arab armies were at the gates of Constantinople, removing the Byzantines from the playing field and leaving the indigenous Berber population as the primary adversary.

For the process of Arabization, 'Uqba's most important act was founding the city of Qayrawān in 670 CE, the first of the *'amṣār* to be founded in North Africa, and the most enduring, the only early Islamic *miṣr* which remains on its original site today. It was founded inland, to keep it safe from Byzantine naval attacks, and like the other *'amṣār*, was originally divided by tribal affiliations, again with Qurayshites and other central Hijazi allies of the Umayyads apparently receiving a relatively more central selection of housing, and the remaining Arab tribes, primarily from Yemen via Egypt, settling together (Ṭāhā, 1989, pp. 61-2). Shortly

after this ʿUqba was recalled, but the commander who replaced him succeeded in confining the Byzantines to Carthage and its environs via a siege and blockade in 678 CE (Kennedy, 2007, p. 212).

ʿUqba returned to the field in 682 CE with a small army. Though this time, he seems to have made it much further, as far as Walila and the Atlantic in the West (further than indicated in Figure 4.1), but this was again more of a raid than a mission of true conquest, and it ended badly with his death near the Roman fort of Tehuda at the hands of the Berber king Kusayla ca. 683 CE. The Arab commander retreated from Qayrawan, though it seems many of the inhabitants remained (as we have accounts that they were treated well by Kusayla, who made it his short-lived capital.) Shortly thereafter, Kusayla retreated into the mountains and the Arabs recaptured Qayrawan without a fight, so the original founding population of Qayrawan seems to have remained largely intact.

The third phase finally consolidated Arab rule enough for the establishment of further Arabic speaking centers. An army was assembled in 74/694 of some 40,000 troops, consisting of Syrian troops and commanded by Ḥassān b. al-Nuʿmān al-Ḡassanī, that is, a member of the pre-Islamic Ghassanid federation (Ṭāhā, 1989, p. 69). These troops were originally stationed in Egypt for a short time, as it was unclear whether they would be needed in the civil war that was ongoing in the Arabian Peninsula, but they then moved into North Africa. Their first task was to eliminate the Byzantine resistance, which they did, finally conquering Carthage and destroying it completely, before founding the nearby city of Tunis in 80/699 as a major naval base from which they could combat Byzantine naval power (Sebag, *El2*, “*Tūnis*”). Tunis was settled largely by members of this army, as well as by a number of Coptic laborers brought for ship building. This final blow against the Byzantines may have led to evacuations of Christian Byzantines out of the

country, as it had in other areas (Abun-Nasr, 1987, p. 31). By the beginning of the seventh century, then, everything from Egypt to Eastern Algerian was brought under Caliphal rule.

Further campaigning was carried out by another commander Musa b. Nusayr, who extended Caliphal rule to what is now Morocco by ca. 710. This ushered in a short-lived era where North Africa formed its own administrative province, under the suzerainty of the Umayyad Caliphs, and with its primary city and capital in Qayrawan. This was an inherently unstable situation, since there was little the central government could do to control restive elements (whether Berber or increasingly independent Arab governors). However, the biggest problem was that this province was essentially an exploitation province and was essentially treated as a source for taxes and slaves which were sent east.

Indeed, the slave trade appears to have been enormous according to early sources. Berber children, especially girls, were captured and exported in great numbers to the Middle East, where they were often bought by the elite at high prices (recall that the first market established at Fustāṭ is said to have been a slave market.) Kennedy (2007, p. 215) notes for example that the Abbasid caliph al-Mansūr (ruled 754–775) was born to a Berber mother captured during the time of ‘Uqba’s campaign. Whereas accounts of the eastern campaigns focused the booty that was acquired in the form of money and goods, accounts of western campaigns focus on slaves. Though the numbers are likely exaggerated, with sources describing raids which netted 200,000 prisoners, it seems quite likely that the slave trade was a major economic incentive for the domination of North Africa.⁷

⁷ One wonders whether the slave trade was being used to repopulate the eastern Mediterranean following the plagues, and whether it suggests that North Africa was, relative to the Middle East, well populated at this time.

In terms of the linguistic impact of this appalling practice, the slave trade would have displaced a number of the native Berber speaking inhabitants of North Africa, and for the young captives, it would likely have led to them acquiring Arabic very early in life. It, along with a number of other practices, also would have served to emphasize the distinction between the Arabic speaking conquerors and the Berber natives, who even if they converted to Islam were discriminated against in various ways. Even when Berbers served alongside Arabs in the military, they were often treated as cannon fodder, being shunted to the front lines (Ṭāhā, 1989, p. 199). Such a division would likely have decreased the permeability of the two speech communities, or at least kept bilingualism only in the direction of Berbers learning Arabic.

The oppressive practices of the Arab rulers led to the outbreak of a major Berber rebellion in 122/739–740, beginning in the area around Tangier. It is notable that this was an essentially Islamic revolution, carried out by Berber Muslims belonging to a Kharijite sect.⁸ At the time of the revolt the main North African army was on an expedition to Sicily. Therefore, the first force that could challenge the rebels were the upper class residents in Qayrawan who fought as cavalry.⁹ In their initial attack on the rebels, they failed dramatically in what was known as the Battle of the Nobles (ca. 740 CE), so named for the number of important Arab aristocrats who died in the defeat.

The first major army sent to quell the rebellion hailed from the Levant, possibly, and was lead by a Qurayshi leader, presumably well connected to the

⁸ A sect of Muslims who didn't believe in the hereditary ascension to the caliphate or the dominance of Arab Muslims over non-Muslims. See Abun-Nasr (1987, pp. 35-39) for an account of the spread of these doctrines to North Africa and their use by the Berber rebellions.

⁹ Cavalry troops were better off than infantry, receiving significantly more of the plunder during campaigns, creating a cycle that increasingly enriched this elite.

Umayyad elite. They arrived in North Africa in 123/741, and entered Qayrawan, where they imposed a fairly onerous military rule that alienated even the Arabs living in the area. This created a split in the army which may have contributed to their eventual failure (Tāhā, 1989, pp. 204-5). They then met up with the remainder of the North African forces camped near Tlemcen. This combined force was defeated by the Berbers near what is now Fes, and routed. The North African soldiers seem to have retreated back east towards Qayrawan, while the Syrian soldiers, numbering perhaps 10,000 at this point, moved north to Ceuta on the straits of Gibraltar. They were painfully besieged by the Berber armies, but were finally rescued by an agreement with the powers in Andalusia, where they were to play a larger political role.

A second army, coming from Egypt, did finally defeat the rebels at the doors of Qayrawan in 742, preserving the major Arabo-Islamic center in North Africa. However, after this, the unity of rule in North Africa was largely broken, with a number of dynasties dividing the region and operating with different levels of independence (see Figure 4.2). Many of these dynasties were Berber, so they would have contributed little to the Arabization of North Africa (and may have reduced it.) On the other hand, other dynasties founded Arabic-speaking cities, such as the Idrisids who founded the city of Fas in 192/808, populated with a contingent of Arabs from Cordoba in 818 and then from Qayrawan between 824 and 826 (Dumper and Stanley, 2006, p. 151; Abun-Nasr, 1987, p. 51).

Similarly, the Aghlabid dynasty which ruled Tunisian semi-independently starting in 800 CE may have led to a faster rate of Islamicization (and possibly Arabization) as a result of their independence from the central power of the caliphate. This dynasty paid tribute to the caliphate, but was ruled internally, making it more of a settlement colony than an exploitation colony (see Section 4.1.2 and

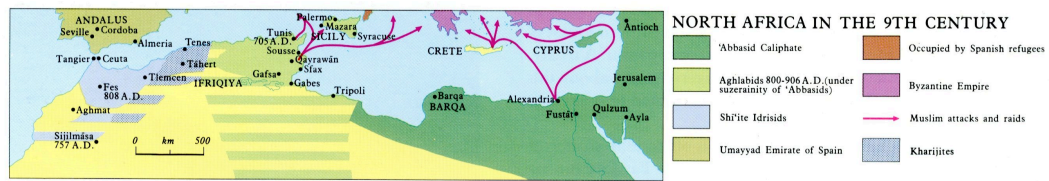


Figure 4.2: North African dominions in the ninth century from Ajaya and Crowder (1985)

Mufwene, 2004; Papaconstantinou, 2012). One result of this independence is that, while Tunisia began to be Islamized later than Egypt, it rapidly caught up with the pace of Egyptian Islamicization, so that the percentage of Muslims in the population were approximately the same in the two countries shortly after conquest (see Figure 4.3).

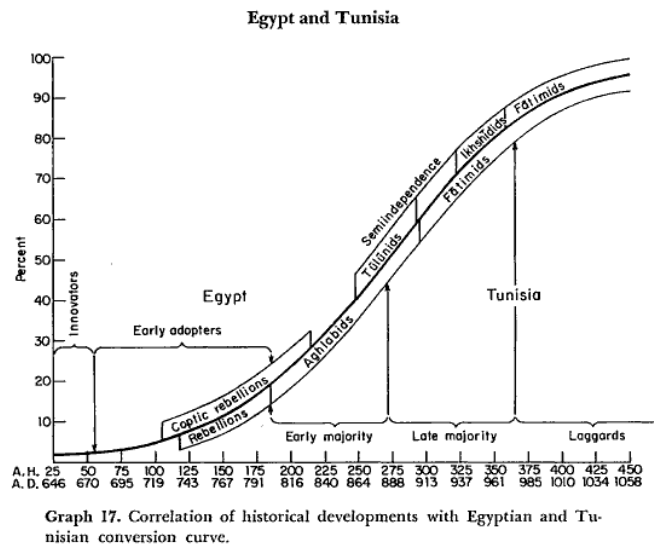


Figure 4.3: The Islamic conversion curves from Egypt and Tunisia showing their parallel development from Bulliet (1979, p. 97)

Thus, the initial Arabization of North Africa seems to have been aided by a

pre-Islamic decline of urbanization and settled populations, which would have reduced somewhat the density of settlement in North Africa. Following this, Arabization would have centered on the major cities in Tunisia, with Qayrawan founded by Hijazi and Yemeni troops who had probably been in Egypt for about 30 years, from the foundation of Fustāṭ in 641 CE until the second campaign into North Africa in 669 CE. Tunis was founded in 705 by a largely Levantine army, which might have included the original Levantine residents from Banu Ghassan, among others. Any other settlements further west probably would have been eliminated by the Berber rebellions in the 740s, with the remaining Arabic speakers confined to Tunisia.

Later developments in the ninth century may have begun Arabizing the western part of North Africa, such as the foundation of Fes in 808 and subsequent migrations from Andalusia and Qayrawan. There may have been some influence from the slave trade, which would have reduced the native Berber speaking population (the slaves, growing up in the Middle East, would not necessarily have participated in any meaningful way in the Arabization of their homeland). The rapid Islamicization of North Africa may also have played a role in Arabization, unlike in Iraq, conversion to Islam occurred normally at the level of the tribe, and not the individual, and so it was probably less disruptive.

Therefore, the evidence appears to be limited for the required settlement patterns that would lead to extensive Arabization of North Africa prior to the tenth century CE. At that time, North Africa was unified under the Fatimid dynasty, which while it had Berber roots, was a dynasty which was ruled from largely Arabic-speaking areas, first Tunisia then Egypt, and which had significant influence over North Africa. However, the Fatimid's role in Arabizing the region has often focused on their role in bringing the Banū Hilāl to North Africa, an event

whose linguistic import may be greatly exaggerated, as we will argue below.

4.2.1 Westward Migrations

The traditional wisdom in Arabic dialectology is that there is a clear distinction between two kinds of dialects in North Africa, what are termed “pre-Hilalian” and “Hilalian” dialects. The distinction between these dialects is very similar to the distinction between sedentary and Bedouin dialects in the eastern Mediterranean – Hilalian dialects tend to be spoken in rural areas, while pre-Hilalian dialects tend to be spoken in urban areas. Hilalian dialects tend to retain interdentalals, and have a /g/ realization of Q variable, much like their counterparts further east.

The traditional viewpoint is presented by Palva (*EALL: “Dialects: Classification”*), though this viewpoint appears to originate in W. Marçais (1938/1956), which in turn seems to be based on the works of the 14th century historian Ibn Khaldoun:

The Western dialects can be divided into two major groups: the so-called pre-Hilālī sedentary dialects and the Bedouin dialects. The former hark back to the first phase of Arab immigration (7th–10th centuries C.E.). The rural dialects of the Jbāla in northern Morocco as well as those spoken around Nedroma in the northwestern corner of Algeria and in the neighborhood of Djidjelli and Collo in northeastern Algeria also belong to this phase. These dialects display considerable substrate influence from Berber languages.

In the 11th century the originally Najdī tribes of Banū Sulaym and Banū Hilāl and the southern Arabian tribe of the Maʿqil moved westward and occupied the North African plains and steppes. At present, Sulaymī Bedouin dialects are spoken in Libya, southern Tunisia, and

northeastern Algeria; eastern Hilālī dialects in central Tunisia and eastern Algeria; central Hilālī in central and southern Algeria; northern Hilālī in the northern part of central Algeria; and Maʿqilī dialects in northwestern Algeria and Morocco. The differences between the Bedouin dialects in the whole Western dialect area are relatively slight. In the Maʿqil and northern Hilālī dialects *j > ž, ġ is retained, and the 3rd pers. sg. masc. suffix pronoun is -ah, whereas the counterparts in the central Hilālī dialects are *j > ž, * ġ > q, and -u.

The long belt of urban pre-Hilālī dialects begins with the old Tunisian cities of al-Qayrawān, Mahdiya, Sousse, and Tunis. In Algeria it continues with the littoral cities of Skikda, Djidjelli, Dellys, Cherchell, and Ténés, and the interior cities of Constantine, Médéa, Blida, and Miliana. In the westernmost part of Algeria the pre-Hilālī dialects include the dialect of Tlemcen, the old urban center of Orania, surrounded by a wide area of Bedouin dialects, and to the northwest of it, the dialect of Nedroma. In Morocco, old urban dialects are spoken in Old Fes, Rabat, Salé, Taza, Tangier, and Tétouan

One can separate this into a number of claims:

- Rural dialects in North Africa are direct descendants of originally Najdi or Southern Arabian dialects
- Rural North African dialects trace their presence to the 11th century
- Certain cities in North Africa were colonized prior to the 11th century and their language reflects that earlier stratum of language

One can add another claim that is seen frequently in the literature (see e.g. Sajīnī, 2007, pp. 129-131), that Banu Hilal were significantly more successful in Arabizing North Africa due to significantly greater numbers, i.e. they caused a demographic change rather than a simply linguistic change:

إذا كان الفتح الاسلامي للمغرب قد حوله إلى بلد مسلم فإن الغزو
الهلالى قد حوله إلى بلد عربي. بالغزوة الهلالية قد أحدثت في
المغرب تعديلا جنسيا فحولته إلى شعب تجري في عروقه الدماء
العربية ...

Whereas the [initial] Islamic conquest of the Maghreb changed it to an Islamic country, then the Hilalian conquest changed it to an Arab one. With the Hilalian conquest a racial change occurred in North Africa, so that the people became one in whose veins flowed Arab blood [...] (ibid., p. 129)

Much of this model of the Arabization of North Africa seems to rest on two main sources. The first is the famous historian Ibn Khaldoun, writing in the late fourteenth century, for whom the Banū Hilāl represent the ravaging pastoralists who conquer declining civilizations, and whose view may therefore be colored significantly by his theoretical approach. The second source seems to be the oral epic of the Banū Hilāl, to which Ibn Khaldoun had access and which later writers rely upon as well, including W. Marçais (1938/1956, p. 186), and which seems to have biased writers towards viewing the migration as essentially that of the Banū Hilāl tribe, though many other tribal groups appear to have participated (for an overview of the oral epic, see Reynolds, 1989).

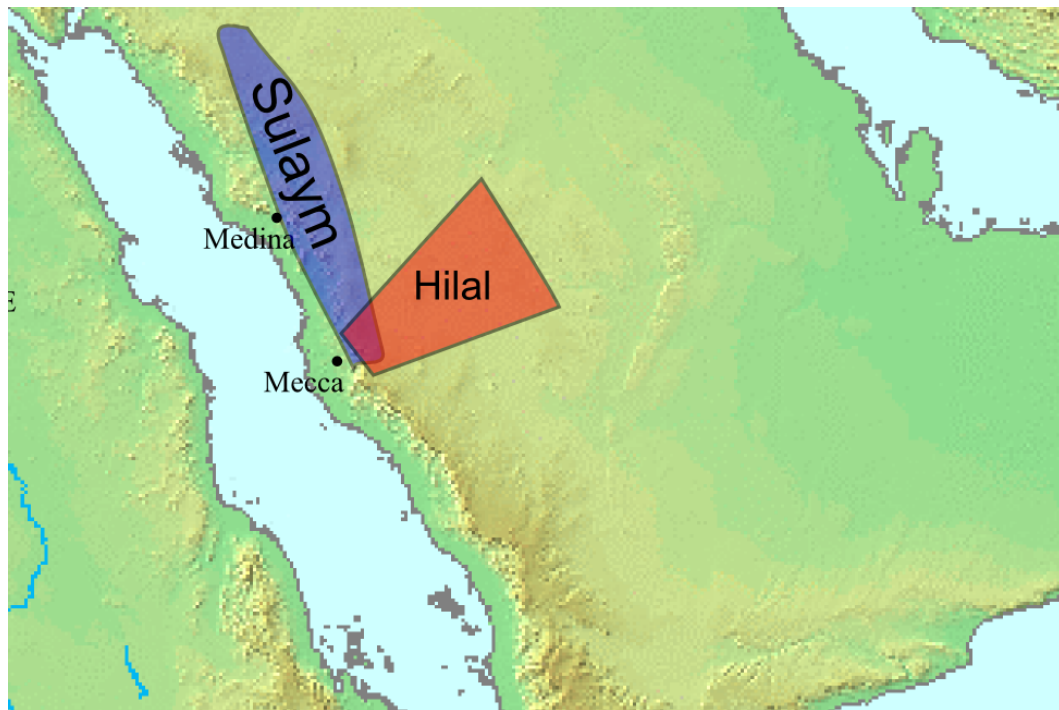


Figure 4.4: The approximate original homelands of Banū Hilāl and Banū Sulaym per the descriptions in Sajīnī (2007, pp. 28, 34 respectively)

4.2.1.1 Historical Overview of Banū Hilāl

What history we do know of the Banū Hilāl creates even more questions about the role they played in Arabizing North Africa.

The two largest tribes that are mentioned in this literature are Banū Hilāl and Banū Sulaym. These groups seem to have come largely from the western part of the Arabian peninsula (see Figure 4.4), and appear to have had close links with Quraysh and the early Islamic community. Note that they were quite mobile, however, with branches in al-Ḥira and Palestine, and Ibn Khaldūn (1377 CE / 2006, Vol.VI, p.16) suggest that their seasonal migrations may have taken them from their Arabian homelands to Syria and Iraq. Banū Sulaym were originally combative towards the Muslims, probably to satisfy their Qurayshi allies, but they were quick to convert and join in the armies of the conquest (Sajīnī, 2007, pp. 30-31). Similarly, Banū Hilāl enjoyed good relations with the Quraysh, including intermarriage pre- and post-Islamically (*ibid.*, p. 36).

The Banū Hilāl and Banū Sulaym were active participants in the Islamic conquests, and their closeness to the Quraysh seems to have allowed them to settle in the Levant, which as discussed above (Section 3.4) tended to be reserved for the Hijazi elite. They are also said to have settled in Iraq in the new cities of Kufa and Basra.

The initial migration of these two tribes to Egypt seems to have been as part of the relocation of Qaysites (the tribal confederation of which they were a part) following the end of the second major Islamic civil war, ca. 109 AH/727 CE. As discussed above in Section 4.1, this seems to have part of a larger power struggle, and the Qaysites were brought into Egypt to balance out the power of the largely Yemeni tribes who had already settled there. However, the two tribes were apparently settled primarily in the Eastern Ḥawf (east of the Nile, NE of Fustāṭ)

and in upper Egypt, with reports of Banū Sulaym living outside of Aswan at the end of the ninth century (Sajīnī, 2007, p. 33).

The second migration, which is seen as the primary migration to Egypt in many of the sources, was the result of these two tribes again taking the losing side in a major civil war, this time in the Qaramite rebellions. The details of the conflict are unimportant here, but in summary, the two tribes were active on the Qaramite side in the theaters of Bahrain and Oman,¹⁰ but they were part of the Qaramite defeat at the hands of the Fatimids in the Levant ca. 368/978. Around this time, the two tribes were exiled by the Fatimids from the Levant into the Eastern Ḥawf and desert, and upper Egypt. Thus, it is some three hundred and fifty years after the initial settlement of these regions in Egypt that the second wave of migrations from these tribes arrived.

The westward migration is said to have been spurred by a series of political machinations, which again may be the influence more of the epic history than the actual history.¹¹ The story goes that a Fatimid client based in Qayrawan, al-Muʿizz b. Bādīs, was asserting his autonomy, and unable to mount a military expedition, a Fatimid official in Egypt decided instead to unleash the hordes of the Banū Hilāl on him.¹² Prior to this, these tribes were apparently forbidden from crossing the Nile westward. In the new policy, not only did the Fatimids allow them to cross the river, but they also granted them money and livestock, encouraging them to conquer further west as part of this punitive expedition.

¹⁰ Hence, we can assume that these tribes represented a peninsular group, even if we are not sure where they were living at the time.

¹¹ Indeed, much of the Arab historiographic tradition is based on anecdotes which can at times feel like a historical novel. Kennedy (2007) makes good use of these stories in his history of the Arab conquests as a tool for exploring attitudes and contemporary cultural assumptions.

¹² In the words of Sajīnī (2007, p. 59), the official sees that the Banū Hilāl and Banū Sulaym's "harm had spread throughout the country" (عم بالبلاد ضررهم).

However, this narrative doesn't entirely seem to be reflected in the facts that follow. If these tribes were to destroy al-Mu'izz's empire in North Africa in some sort of lightening strike, they did not received the memo, and instead slowly moved out into Libya. The tribes are said to have left Egypt sometimes between 440–443/1048–1051 depending on the source, with Sajinī (2007, p. 60) favoring 442–443/1050–1051. Even with a later start date, it is nearly three years before these tribes are said to have conquered Tripoli, though they are said to have started earlier with a conquest of the relatively nearby Barqa in Libya. If they were intended to humble al-Mu'azz, they certainly took their time doing it.

Instead, whatever the original impetus of the migrations, they were rather slow, and resemble a general migration in search of greener fields more than a military expedition. We also have reports of multiple waves of migrations, with the first wave encouraged by the Fatimids as explained above. However, a second wave seems to have followed shortly afterward, lured by stories of the fertile new country, and the Fatimids are said to have exploited this to their fiscal advantage, charging them to cross the Nile in a bid to recoup their original expenses from the first expedition, hardly a way to treat one's reinforcements for a military campaign. Indeed, the military acts of the Banū Hilāl and Banū Sulaym seem to be quite independent of the Fatimid empire, and they quickly became a power in their own right within North Africa.

The tribal confederations did eventually sack Qayrawan and dethrone al-Mu'izz in 449/1054-55, at least half a decade after they first moved into North Africa, hardly a rapid advance in comparison with 'Uqba's one year campaign that covered the same territory while in conflict with a major world empire.¹³ Var-

¹³ The Encyclopedia of Islam (Schleife, *EL2*, "*Hilāl*") suggests that this was a blow to the city of Qayrawan "from which it never recovered", which may have caused a dispersion of Arabic speak-

ious incidents of internecine warfare seem to have spread members of these tribal coalitions around, with the Riyāḥ tribe driving the ʿAdī tribe further west from Tunisia after a battle in 491/1097-8, though both were subgroups of the Banū Hilāl. In the late 540s/early 1150s, they suffered a series of military defeats at the hands of the Almohads coming from Morocco. After this, in the words of Schleife (*EI2*, “*Hilāl*”), “from the 7th/13th century it becomes more and more difficult to distinguish the Hilāl from the Sulaym and other nomadic Arab tribes who followed them and pursued their work of devastation.”

4.2.1.2 Re-evaluating Origins

The traditional narrative, that the second phase of Arabization brought with it a straightforward 5th/11th century example of Najdi Arabic, is extremely attractive. It would give us a direct insight into a dateable, geographically specific set of dialects. However, it would appear to be too good to be true.

The Banū Hilāl and Banū Sulaym first move into Egypt ca. 368/978. It was some seventy years later, ca. 442/1050, that they and numerous other groups were to move into North Africa. Linguistically, this period of time spent in Eastern and Upper Egypt would probably have lead to adaptation of the new immigrants to the already dominant varieties, which in turn likely represented something of a northern Hijazi/Levantine variety of Arabic. Indeed, one of the frequent points of the Banū Hilāl mythos is that they became quite numerous during this short period in upper Egypt (Sajīnī, 2007, p. 40), which if true, would suggest that most of the members of the group that moved into North Africa were born in Egypt. Barring serious segregation of the newcomers and the older residents, it seems quite

ers into surrounding areas. However, the Encyclopedia of Islam tends to be fond of hyperbole, so I am wary of this judgment.

likely that the linguistic variety of the fifth/eleventh century westward migrations reflected whatever Upper Egyptian Arabic was present at this time, more than the originally Najdi.¹⁴ The irony is that the Upper Egyptian Arabic of the time probably reflected that of the earlier migration of Qaysites in the second/eighth century, though of course it would be difficult to tell how much continuity or change there would have been in their dialects.

Moreover, the specific list of which tribes seem to have moved west is quite extensive, though they were apparently under the banner and leadership of the Banū Hilāl. The list seems to include tribes with a wide variety of origins (see *ibid.*, pp. 40-41), including Yemen, though many of the tribes are said to have been Qaysites, a designation which at this juncture was probably only nominally meaningful.

Therefore, the groups of Arabic speakers that moved into North Africa in the 5th/11th century probably represented a mix of dialects, with the majority reflecting Upper Egypt dialects, but possibly some influence from the western Najdi dialects of the 4th century. Most importantly, it is quite unclear what exactly the origins of their dialects were, in contrast to the traditional narrative where the answer is overly simple.

4.2.1.3 Re-evaluating Timing

The most unconvincing aspect of the traditional narrative is the supposed timing of events, especially if we look at the cities which are typically said to speak pre-Hilalian dialects. The cities for which we have convincing foundation

¹⁴ Moreover, Sajīnī (2007, pp. 40-41) suggests that numerous other groups already resident in the areas settled by these newcomers became part of Banū Hilāl confederation, operating as subtribes of the Banū Hilāl.

dates prior to the 5th/11th century include Qayrawan, Tunis and Fes, and these do indeed have dialects which are more similar to one another than they are to nearby rural and nomadic dialects. However, many of the other cities which make the list of those speaking pre-Hilalian dialects appear to have been largely populated after the 5th/11th century. The city of Cherchell in north central Algeria on the coast (with their dialect represented in this study by Jacques. Grand'Henry, 1972) though with older origins, is reported to be essentially an unpopulated or sparsely populated ruin in the geography of Al-Bakrī, composed ca. 460/1068. In the following century, with plenty of time for the Banū Hilāl to have reached this area, it was said to be small but well populated, with nearby Bedouin living off of agriculture and cattle rearing (Yver and Sari, EI2, "*Sharshal*").

Similarly, the city of Djidjelli, on the coast of NE Algeria (with their dialect represented in this study by P. Marçais, 1956), though also having a long history, was controlled by Berber powers for centuries. It may have been raised and repopulated by Al-Bakrī's time, so whatever population lived there was likely relatively recent and small when Banū Hilāl and its allies came to North Africa (Yver, EI2, "*Djidjelli*"). The city of Tlemcen, in NW Algeria (with their dialect represented in this study by W. Marçais, 1902) appears to have been largely ruled by Berber dynasties, and the present town even appears to be a merger of a Berber *miṣr*-type settlement (named Tākrārt, meaning 'camp') and the original settlement, with the towns being merged and walled in 540/1145. All of the names of this town appear to be in Berber, suggesting at least Berber language dominance at all these phases, if not predominance.

These examples suggest that the hometowns of supposedly "pre-Hilalian" dialects date their founding and Arabization as late as the 5th/11th centuries, and in many of these cases there was ample opportunity for members of the tribal

federations that moved to these areas to take part in Arabizing these towns. The only cities which could conceivably show a clear pre-Hilalian stratum of language are those in Tunisia and Libya, the latter of which seem to be heavily Bedouinized today.

Another example is that of Rabat, on the Atlantic coast in Morocco, which seems not to have been of great importance until the 5th/11th century, when it was used as a launching point for the Almohad Berber dynasties' attacks on Spain. In fact, many of the Arab tribesmen defeated by the Almohads in 1153 CE are said to have been settled in the Moroccan coast between what is now Rabat and Casablanca (Abun-Nasr, 1987, p. 93), which suggests that Arabization of this city may indeed have been at the hands of the Banū Hilāl themselves, in spite of the modern literature which considers Rabat to have a pre-Hilalian dialect.

Another aspect of timing is the continued migrations in a number of different directions. There are numerous historical examples of movements from west to east, even among the Arab tribes, and Banū Sulaym are said to have moved back into Egypt in a number of waves over a period of centuries, with a particularly large migration in the 18th century CE when they settled in Western Egypt (Sajīnī, 2007, p. 69). Schleife (*El2*, "Hilāl") has a very confused chronology of migrations, but as quoted above, notes that it is quite difficult to distinguish the 'Banū Hilāl' migrations from others westward migrations.

4.2.1.4 Re-evaluating Impact

The demographic and linguistic impacts of the 5th/11th century migrations seems to be rather exaggerated in the sources. As illustrated above, many of the supposedly pre-Hilalian dialects seem to have been spoken in cities established around this time, so either the actual Hilalian dialects were much closer to what

we now call pre-Hilalian dialects, or the Banū Hilāl simply had very little influence on the varieties spoken in cities, which e.g. in the case of Rabat or small cities like Cherchell seems unlikely. Furthermore, the primary areas settled by the Banū Hilāl and Banū Sulaym, namely Tunisia and northern Algeria, were exactly those areas which were probably the most heavily Arabized even before they came, so their impact would probably have been incremental rather than revolutionary.

The demographic narrative, that Arabs arrived in North Africa in such numbers that they largely overwhelmed the native population, is not supported by the genetic data. In a study of the genetics of Moroccan and the isolated Algerian Mozabites, Bosch et al. (2000) found that populations were genetically quite homogenous, and argue that this evidence means that:

Given that the linguistic and cultural differences among Arabic and Berber speakers is not reflected by their genes, it is a plausible argument that the Arabisation of NW Africa was only a cultural phenomenon with subsequent little genetic impact.¹⁵

Gaibar et al. (2011) similarly found very little affiliation between either geography (Tunisia vs. Morocco) or language (Arabic vs. Berber) and genetics, such that these populations were largely identical. Gérard et al. (2006) were able to differentiate the two geographical areas somewhat, but give the not-unexpected result that the most heavily Arabized area genetically speaking is located in Tunisia and coastal Algeria (see Figure 4.5). Even in an analysis of two nearby villages

¹⁵ Nebel et al. (2002), in a letter to the editor of the American Journal of Human Genetics, argue against this interpretation, on the basis of evidence that links Yemeni, Northern Palestinian, and some North African populations. This correlates nicely with our evidence of migration generally, but does not appear to show the kind of overwhelming population movement normally ascribed to the Banū Hilāl and other tribes.

(separated by some 40km), one said to be pre-Hilalian Berber, and the other to be Arab, the genetic markers in both villages were more similar to other Berber populations than Arab populations, and the general genetic environment is described as a “patchwork” which defies easy analysis (Cherni et al., 2005).

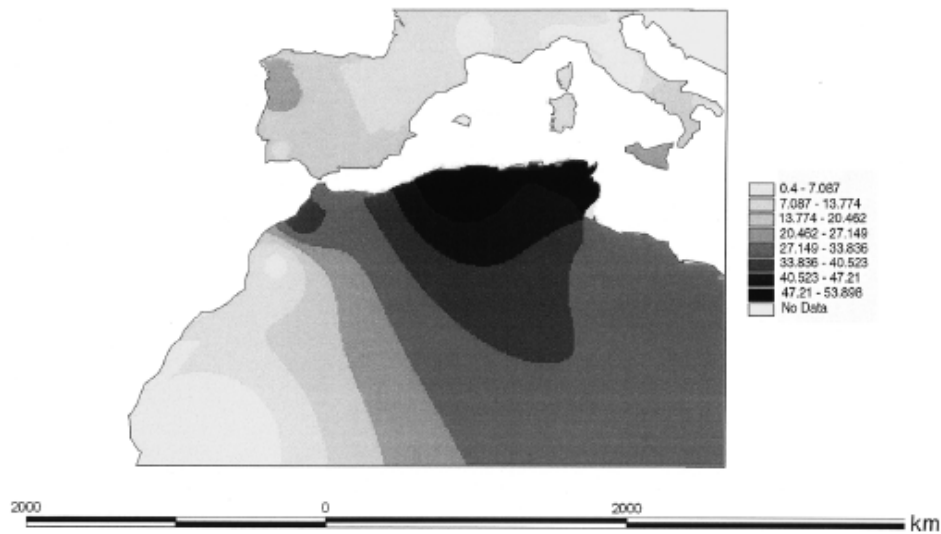


Figure 3. Isofrequency map of subhaplotype Va. The resulting values were divided into eight equal classes.

Figure 4.5: A diagram of Arab gene markers in North Africa, produced through interpolation (not from extensive sampling), from Gérard et al. (2006).

Finally, the traditional sources also propose an east-west division of North Africa by the Banū Hilāl and Banū Sulaym, with the former taking Tunisia and westward, and the latter taking possession of coastal Libya. If indeed these two tribes’ dialects which had diverged as far back as the pre-Islamic era, as per the narrative, we’d expect to find a much stronger east-west dialect division than we currently have in North Africa. Indeed, even if these dialects were quite similar (e.g. both had assimilated to 10th century Upper Egyptian Arabic, as is likely), a thou-

sand years of differentiation would likely look more extreme than what we have (refer again to Palva, *EALL: “Dialects: Classification”*, quoted above). Instead, the evidence suggests more recent replacement, which would be consistent with continued migrations into sparsely populated areas, in conformity with the Bedouin Paradox.

4.2.1.5 Western Migrations Reconsidered

Given the evidence presented above, we must consider an alternative narrative of westward expansion. Whatever effect the tribal migrations of the 5th/11th century had has been greatly exaggerated, and does not seem to have been the watershed moment that changed the linguistic situation in North Africa. Instead, it is more convincing to posit a longer term, continuous series of migrations from east to west bringing Arabic speakers into North Africa, and creating a linguistic division between older populations and new arrivals. Indeed, east to west migrations seem to be quite common in the history of North Africa, as such migrations are reported even in pre-Islamic times Kennedy (2007, p. 205).

However, given the evidence that the shift to Arabic was cultural rather than genetic, and that Arabs constituted a minority population in an area often ruled by Berber dynasties, we need some mechanism to account for the Arabization of North Africa at all. This seems to have been accomplished by some forms of cultural pressures, particularly discrimination of Arabs against Berbers, discussed above, which may have led to the Arabization of Berbers as a means to escape this discrimination. Ibn Khaldoun gives an account of cultural assimilation by Berber tribes to Arab lifestyles (Sajīnī, 2007, p. 130), and there are various accounts of Berber tribal federations taking Arab geneologies, probably in order to avoid discrimination and have greater cultural mobility (Sajīnī, 2007, p. 130; Nebel et al.,

2002). These incidents appear to represent a cultural pressure for assimilation to Arabic cultural norms, which may have been a precursor to linguistic Arabization. Nonetheless, the process of shift to Arabic, which in much of North Africa is still incomplete, was likely very slow and cannot be traced specifically to the Banū Hilāl migrations.

4.2.2 Effects of the Reconquista on North Africa

The Reconquista, that is the (re)conquest of Muslim Andalusia by Christian forces, began in the eleventh century, and can be said to have continued until the fall of Grenada in 1492, though expulsions of Muslims continued for some time after this. To go into any detail about this series of events would require a much more comprehensive treatment than is possible here, but we will try to have a basic outline of how refugees from the reconquest might have settled in North Africa and contributed to the Arabization of that region.

al-Ḥamrūnī (2009, pp. 4-5) divides the reconquista into four major phases. The first is the initial reclamation of the north and reorganization of the kingdoms there that lasted for the first several centuries of Muslim rule. The second was the fall of Toledo and Teruel in 1085 and 1171 respectively, a phase in which Muslims living in the newly conquered lands were basically tolerated. The third phase continued with the conquests in 1275, which brought the Christians to the doors of Granada, which was to remain essentially a stalemate until the final phase, the conquest of Granada in 1492. The third phase onward is when Muslims began to be expelled, or to emigrate in earnest to Muslim lands, whether in the south of Iberia or across the Mediterranean and Atlantic, though the various treaties that were enacted tended to vary as to whether they allow Muslims to remain or not. The final blow for Islam came in 1609, when the Christian powers began forcibly

expelling Muslims from Iberia.

Though the population estimates vary, there was likely a huge change in the Muslim population in Spain, whether from reconversion back to Christianity by groups which have previously converted to Islam, or by migration and movement out of the peninsula. Harvey (1990, pp. 7-9) finds Glick's (1979) estimate of a total Muslim population of 5.6 million by 1100 to be perhaps exaggerated, but nonetheless, there was a significant decline in Muslim population to about 1 million by 1300.

Whatever the numbers, that drop in population is of particular interest to this study of the movement of Arabic speakers from Andalusia to North Africa. However, the majority of the secondary research is on the later periods, starting at earliest in about 1200 CE (*ibid.*), but with most works focusing on the period starting in the sixteenth when official decrees of expulsion of Muslims were promulgated (al-Ḥamrūnī, 2009; Harvey, 2005; Razzūq, 1989) following the fall of Muslim rule in the Iberian Peninsula.

Though the Reconquista was the major driving force in much of the Muslim migration back to North Africa, there was frequent movement between the two areas when they were united by Muslim rule. The two areas were linked by trade, and movement between them should have been relatively easy. Even as late as the 1440s, we have reports that refugees from Andalusia moved to Fes, but found the conditions there intolerable, and moved back to Andalusia (Razzūq, 1989, p. 146).

Many of our reports of migration are of the movement of entire groups, though that may be an artifact of the historiographical tradition that would only report on movements of that size. Hence, we hear that the Banū Ašqilūla, a major power in thirteenth century Andalusian politics, moved back to North Africa en masse in 1288 (Harvey, 1990, p. 160). When the mountainous Muslim stronghold

of Ronda in the far south fell in 1485, the terms of the surrender allowed for safe passage to North Africa, and the Muslim inhabitants of the city appear to have moved at once (Harvey, 1990, p. 286). Such movements of large, intact groups, would likely have a stronger linguistic impact on the host countries than a trickle of individual migrations

There is also a class dimension — the earliest migrations to North Africa were likely carried out by members of the wealthier classes whose wealth would have been in moveable goods rather than land, and who could afford the longer journeys to North Africa. Poorer people could not make so long a journey, and so throughout the initial phase of the reconquista many Arabs seem to have moved just as far as necessary into territory still under Arab rule. After 1275, when most of the peninsula was under Christian rule, this movement would have largely centered on Granada which was to be the major Muslim stronghold until the seventeenth century (*ibid.*, pp. 51-53).

We do know where major groups of Andalusí Arabs settled: Many of the refugees settled in northern Morocco, near Tangiers, Tetouan, and Oran, and this area was also the unloading point for the Valencian refugees expelled in 1609, who totaled nearly 120,000 persons (Harvey, 2005, pp. 316, 360). Fes was a major settlement point even before the Reconquista, with a “Mosque of the Andalusians” founded in 245/859, and with continued settlement in a specifically Andalusian quarter. Others settled in Chefchaouen also in northern Moroccan, which was only established around 876/1471. The newer half of Rabat-Salé, now the capital of Morocco, appears to have been populated largely by Andalusian in the 1600s CE. Refugees also settled further west, including Tunisia, where after the final expulsions in the 1600s, many Andalusians were given tax incentives by the Ottoman empire (*ibid.*, pp. 358-9). The supposedly pre-Hilalian populated town of

Tlencen also appears to have absorbed its share of Andalusian refugees (Razzūq, 1989, pp. 129,155).

In all, the total number of Muslim moriscos expelled in the sixteenth century numbered perhaps as much as half a million, out of a total population in Iberia of some eight million. On the other hand, this is said to represent a “community long in decline” (Harvey, 2005, pp. 12-13), which may well have declined both through conversion to Christianity, as well as through migration to North Africa, possibly in much greater numbers, as discussed above.

Jews were also expelled from Andalusia, though the chronology differs in some respects from the Arab migrations. In contrast to the Arabs, the reconquests of the 11th through 13th centuries did relatively little to alter the lives of Jews in the newly Christian lands, and in fact there seems to have been some migration from Muslim to Christian lands at this time (Ashtor et al., *Spain*, pp. 71-72). However, in the late 14th century, starting in 1378, a campaign of repression began in and around Seville, and this spread throughout the Christian lands. Many Jews were massacred, forcibly converted, and some began to escape to North Africa.

Before this time, Jews had faced oppression in Morocco and southern Andalusia at the hands of the Almohads, whose strict religious stances left little room for non-Muslims. However, the Merinids who followed them beginning in 1269 had significantly more liberal policies towards Jews, making Morocco and North Africa a more attractive destination when the policies in Andalusia turned against them (Corcos, H. J. Cohen, and Laskier, *Morocco*). The primary destination for these Jews leaving after the late fourteenth century was Algeria, where emigrating Jews from Catalonia and the Balearic Islands had commercial ties, and where they thought they would find a more welcoming home. There was of course some tension between the newcomers and the host community, just as with Arab mi-

grants, with the incomers having greater financial resources generally (Corcos, Saadoun, and Attal, *Enc. Judaica*, “Algeria”).

The final expulsion of Jews from Andalusia came in 1492, though they had been expelled from various other parts of the peninsula earlier than this date. With the research conducted here, there seems to be no detailed account of where these exiles fled to, many seem to have joined earlier diaspora communities in North Africa, or further east within the Ottoman empire whether in Anatolia or the Levant (Ashtor et al., *Spain*).

In summary, the population movements as a result of the reconquista probably began as early as the mid-eleventh century, though they would have accelerated in the thirteenth as expulsion policies of Muslims and Jews became more strict in the reconquered Christian lands, and the final major migration of Jews would have been in 1492, with Muslims getting expelled finally starting in 1609 and continuing for some time after this. The fact that many of these emigrants settled in towns where are said to have “pre-Hilalian” dialects brings up the speculative possibility that the dialects of these cities are simply heavily influenced by the Andalusian dialect rather than representing a much earlier linguistic period. However, this is necessarily speculative given the lack of better data on the size and direction of the migrations.

4.3 Andalusia

This section will be as brief as possible, since Andalusian Arabic is poorly attested in our sources (primarily in its literary or partially elevated varieties) and because it only constitutes a single datapoint in our study here. Moreover, the history of the Arabization (including whether Andalusia was fully Arabized at all) is extremely contentious, and beyond the scope of this study (see the comprehensive

summary of positions in Zwartjes, *EALL : “Andalus”*). We will focus here primarily on the question of who settled Andalusia, and their linguistic contributions to the Arabic of the Iberian Peninsula. We will also briefly touch upon the question of how extensive Arabization was within Arab-controlled lands.

The initial conquest of Andalusia was at the hands of a small, largely Berber army (at most about 10,000 strong) under the command of a Berber Muslim, Ṭāriq bin Ziyād,¹⁶ who took advantage of a civil war among the Visigoth rulers of Iberia to rapidly conquer as far as Toledo in the center of the peninsula in 92/710-711 CE. The following year, a much larger, primarily Arab (with some Berbers as well) army was sent by the governor of North Africa. This army numbered approximately 18,000 men, and at first worked to consolidate Ṭāriq’s earlier gains in the south. The army went on to conquer most of the peninsula by about 715, though as in many other places, they failed to hold areas above the 1000 elevation meter line (Kennedy, 2007, pp. 313-318). The Arab armies continued to raid even further north, until a defeat near Tours or Poitiers in north central France in 732, which is largely seen as marking the furthest European penetration of Arab armies (*ibid.*, p. 320).

The second army had been largely recruited from forces already present in North Africa, particularly in Qayrawan, though they are often spoken of as Yemeni with a primarily Qurayshi leadership (Ṭāhā, 1989, p. 94). Since Qayrawan was founded in 670 CE, and this army moved into Andalusia around 712 CE, this gives us about a forty year lag time, or possibly two generations of military aged men, so their dialect may well have represented a moderately stable Qayrawan koine at this date.¹⁷ The army was apparently assembled with the goal of gathering the

¹⁶ The namesake of Gibraltar < Ar. *ḡabal ṭāriq* ‘Ṭāriq’s mountain.’

¹⁷ The inputs to this variety having been both Hijazi and Yemen, possibly with some influence from

booty available in Iberia, and this may explain the presence of many elites in the army, who would have been eager to take exploit their positions for further gains.

It is these troops who seem to have been the initial settling group of Arabic speakers in Andalusia, with many remaining in the areas that they had conquered (Ṭāhā, 1989, p. 100). The total size of this group, including dependents and others was likely between 25,000–30,000. Though, unlike in other conquered territories, few *ʿamṣār*-type settlements were established, this initial group seems to have settled with Berbers moving into rural areas, and Arabs settling in the cities, though there was much variation (Hernández, 1998, p. 61).¹⁸ Though Yemeni is a somewhat vague term, we do have some sense that they may have originally hailed from eastern Yemen, since we are told that quite a few members of the Hadramawt tribe were settled near Seville.

Another wave of settlers came in the form of the defeated army which had originally been tasked with breaking the Berber rebellions. They had succeeded in defending Qayrawan, but their campaign further west failed and they were blockaded in Ceuta by Berber forces. Eventually they were able to negotiate a rescue into Spain, on the condition that they aid in crushing the Berber rebellions in Iberia, then return to the Levant. This army, which had numbered some 30,000 at its outset probably largely drawn from the Levant, was reduced to about 10,000 ragged, starving soldiers by the time they arrived in Andalusia in about 124/741. However, they recovered enough to help subdue the rebellion, and in spite of the agreements they had made to leave the country, ended up settling in Andalusia, particularly in the countryside.

the generation or two that they stayed in Egypt.

¹⁸ An extremely detailed list of settlements is available in Ṭāhā (1989, pp. 118-130), but given how little linguistic data we have for Andalusia, this is not worth the effort to follow up on more thoroughly.

There was a great deal of resentment from the original settlers against the newcomers, including several military actions in which the original settlers attacked but were defeated by the Syrians, who may have benefited from a closer social cohesion than their longer resident opponents (Hernández, 1998). Though these newcomers were at first poor, they linked themselves to the rise of Umayyad state-in-exile after the arrival of the Umayyad scion ‘Abd al-Raḥmān I in 756 CE fleeing from the Abbasids in the east.¹⁹ Even before this time, the newcomers were exempt from taxes imposed on the earlier settlers (*ibid.*, p. 134) and their privileges increased under the Umayyads, for whom they represented a significant portion of their powerbase. Other groups that had been in the Levant also seem to have fled to Andalusia around this time, so that much of this settlement can be said to have been Levantine, or at least settled in the Levant for several generations (*ibid.*, pp. 149-150). This group of Levantine extraction may have been the primary Arabizing force in the country, as Ṭāhā (*ibid.*, p. 219) speculates that Syrians may have been deliberately settled in Andalusia to shore up an Arab presence against the Berber and indigenous populations.

The Berber populations may well have been numerous, but they were treated poorly by the Arabs, as they had been in North Africa. While they played a major role in conquering the country, they were rarely appointed to any positions of leadership, and they were generally relegated to poorer places (*ibid.*, p. 207). These factors seem to have led to a Berber rebellion in Andalusia around 741 CE, which was defeated by a combined force that included the newly arrived Syrians, probably further weakening the power of Berber to the benefit of the Arabs.

¹⁹ Though the break between supporters and opponents of the Umayyads did not break exactly along the fault between new and old settlers, such that he was said to have been supported by early “Yemeni” settlers and opposed by settlers hailing from Damascus (Ṭāhā, 1989, p. 238).

In summary, the groups which settled Andalusia came in two phases: The first was a limited numbers of settlers from the Arab armies in North Africa, particularly settlers from the Hijaz and Yemen, including elements from south-eastern Yemen, as well as Berber settlers who were probably a numerical majority, but politically and socially marginalized. This was followed by a larger influx of Arabic speaking settlers from the Levant, who had greater political power and various advantages over the previous wave of settlers.

As mentioned above, these settlers did not form *'amṣār* style garrison communities. With no *diwān* system of payment, there was no need to create concentrated urban settlements, and so people tended to live wherever convenient after the conquests. Only a small number of cities were founded or further fortified (such as Madrid), with many simply continuing or growing (such as Seville and Cordoba), and a number moving a small distance, particularly in the 8-10th centuries CE (Kennedy, 1998b). There were of course a number of towns that were also abandoned. Nonetheless, the Arabs are said to have moved into a land where “cities were small and comparatively undeveloped and [...] much of the countryside seems to have been very sparsely populated.” (Kennedy, 2007, p. 309) The relatively sparsely populated country and the (linguistically disruptive?) small shifts in site of some cities may have contributed to whatever success Arabic had.

As alluded to above, the question of the extent of Arabization is a complex one, and one which is difficult to assess from the evidence available. There is at least one text complaining about the loss of Latin or Romance by the mid-ninth century CE, from a writer in the Muslim capital of Cordoba, which is not entirely unexpected given the location (Wasserstein, 1998, p. 3). Moreover, the writer appears to be discussing highly literate uses of language, not everyday speech, and

we have little clear evidence regarding that level of communication.²⁰ Moreover, the shifting nature of the political and demographic situation in Andalusia, where from the 11th century increasing swatches of territory were placed under Christian rule, makes it hard to make any firm statements on language use in the Iberian peninsula as a whole.²¹

The summary of Wasserstein (1998) seems like a reasonable scenario: Different groups would have changed language at different rates, with convergence to either bilingualism between Arabic and Romance (with different degrees of dominance) or possibly Arabic monolingualism in the cities in Muslims areas. He suggests that by the eleventh century, “Latin was dying out in all of its forms” within Andalusia, though it is not clear exactly what the boundaries are of this area. Nonetheless, the conclusion can be drawn that Arabic was a well-established linguistic variety used in daily life in some parts of Andalusia. That Arabic variety would have been derived from the speech of the original inhabitants of Andalusia, who had been 3rd or 4th generation North African Arabs, and later immigrants from the Levant. There would have been extensive periods of bilingualism with Romance and possibly Berber languages.²²

²⁰ Zwartjes (*EALL*: “*Andalus*”) quotes Federico Corriente as saying, in a mass of contradiction, that Romance became: “the prestigeless language of women, peasants, and slaves. Most people still understood it to some degree but very few cared to speak it any more, although it remained in use in images of domestic scenes.” One can only imagine the difficulties of domestic life when only women spoke Romance while men spoke Arabic!

²¹ Indeed, one wonders how many of the contradictions in the sources surveyed by Zwartjes (*ibid.*) for the Arabization of Andalusia could be resolved simply by greater temporal and geographical precision.

²² If Romance had penetrated fairly far in North Africa, it is also possible that the ethnically Berber troops that partially conquered Andalusia actually spoke Romance, and so there would really have been more of a distinction between Romance and non-Romance (i.e. Arabic) speakers.

4.4 Peripheral regions

There are a number of peripheral regions which speak Arabic but which are generally outside of the center of the Arabic speaking world. The fact that many of these areas no longer participate in the use of literary Arabic, and hence diglossia, makes them appealing to researchers who view them as pristine, dateable examples of earlier states of Arabic (Al-Jallad, 2009; Owens, 2006). There is no doubt that these dialects can be helpful, but they only represent a small number of the sample points in this study, and so the study of their history will be rather more prefuntory than the regional histories above.

4.4.1 Malta and Sicily

The linguistic history of Sicily is not of great importance in and of itself, since it is poorly documented and no living dialect currently exists in Sicily. Instead, its importance lies in the fact that it was likely the launching point for movements into Malta.

Muslim raids on Sicily, originating in Tunisia, began as early as 32/652, but these were essentially petty raids rather than invasions, and it wasn't until 212/827 that Arab forces began a concerted invasion effort with some 10,000 troops who were drawn from North Africa, followed by further reinforcements in 215/830 (Grassi, *EL2*, “*Şiḳilliya*”). This conquest of the island was completed by 289/902, at which point Muslim colonists were encouraged to settle in the newly conquered areas. The settlement accelerated in the following two decades, though Arab Muslims primary colonized the northeast of the island, while Berber speakers colonized the south, and Romance speaking Christians probably remained in the northwest (Agius, 1996, pp. 26-8). Sicily was basically a province of whatever power was in control of Tunisia, but following the move of the Fatimids to Egypt

in 362/973, Sicily became relatively more autonomous.

Sicily flourished as a Muslim state for the next sixty years or so, but began to break down into competing principalities around 431/1040. Shortly after this, in 1061, Norman troops began to land on the island, and it was entirely under their control by 1091. Nonetheless, Norman rule was very tolerant of the Muslims, integrating them into the army and even discouraging them from conversion to Christianity, and Arabic seems to have continued to be spoken even at this time (Grassi, *EL2*, “*Şikilliya*”). The Normans also moved into North Africa, capturing much of the Tunisian coast between 1135-1153 (Agius, 1996, p. 35), which likely promoted movement between the two areas. The Normans were eventually replaced by German Swabians, who took Sicily between 1194–1197. It was in 1197 that Muslim riots against conditions in Sicily began in earnest, resulting in a rebellious army which fought until 1223, when they were defeated and Muslims living in Sicily were exiled to a colony in Lucera, Italy (*ibid.*, p. 37).

The conquest of Malta is said to have begun in 256/870, though there may have been basic raids earlier. The forces that invaded Malta at this time appear to have been drawn from Sicily, where they could have represented second generation residents of Sicily following the initial invasions (Brincat, 2011). However, Malta is said in some sources to have been basically ruined and sparsely inhabited after that time, and it wasn't until a second invasion, also from Sicily in 1048-49, that it seems to have been repopulated by Arabs, an idea which Brincat (*ibid.*, pp. 35-49) finds support for from the historical and linguistic evidence.²³ If this

²³ On the other hand, if 1048 is the date when Malta is Arabized, it's very hard to explain the robustness of Maltese vis-a-vis a Norman invasion only a decade later, though the Normans did tend to have liberal policies towards Muslims.

interpretation of the historical accounts is correct, i.e. Malta was depopulated and then repopulated, whatever earlier linguistic history would not be important to understanding the development of Arabic in Malta, especially since it is still unclear what language would have been spoken in Malta if it had been inhabited (Brincat, 2011, chapter 1).

Malta was then conquered by the Normans in 483/1090, though Muslims continued to be allowed to live on the island until 647/1249, where they were forced to either convert, or be expelled.

The influence of the new European conquerers appears to have been extremely limited at first, with Maltese acting as a “as a secret speech which excluded foreigners (Luttrell, 1991, p. 39).”²⁴ This seems to have remained the case as long as Malta was treated as the outlying province by various European powers,²⁵ but a major shift in language appears to have occurred following the transfer of Malta by Charles I of Spain to the Order of the Knights of St. John of Jerusalem in 1530, the first time the ruling group had actually been based on the island and integrated itself into the native population via permanent settlement (Mifsud, 1995, 24-26ff). This seems to have been the catalyst for the beginning of a much more extensive period of influence of Romance into Maltese, suggesting a “tipping point” scenario whereby change occurred quite slowly until an event or series of events causes much more rapid linguistic change to occur.²⁶

²⁴ Cited by Cremona (1994, p. 284), an excellent example of what Ross (1997, p. 239) refers to as ‘esoterogeny’, the deliberate attempt to exclude a linguistic outgroup.

²⁵ Again, supporting the distinction drawn by Mufwene (2004) between settlement and exploitation colonies.

²⁶ For more on this, see Dixon (1997) (but also Bowerman (2006)) for the idea of “punctuated equilibrium”, though this idea of a tipping point is somewhat different. The question of whether it was simply the beginning of a more local form of rule that caused the linguistic change, or other events (such as the Siege of Malta in 1565, where nearly a third of the inhabitants were killed) bears further investigation.

Maltese and Sicilian Arabic seem to have diverged in the eleventh and twelfth centuries, with Sicilian Arabic showing significantly more influence from Romance-Arabic bilingualism than Maltese Arabic. Though both show phonological changes probably due to Romance influence, Maltese has a rather different phonological inventory than Sicilian. Similarly, a number of nouns in Sicilian Arabic take gender in accordance with a Romance substrate, rather than the predictable gender patterns typically seen in Arabic (all feminine nouns end in the morpheme /-a(t)/ with some exceptions), so that *sayf* ‘sword’, masculine in Classical Arabic and most dialects, becomes feminine in Siculo-Arabic on analogy with Romance, e.g. Italian ‘spada’, Spanish ‘espada’ and French ‘épée’, all of which are feminine (Agius, 1996, p. 145). In Maltese, the gender of this and other words appears to remain as in other Arabic dialects.

Maltese, then, appears to represent a dialect which was drawn from Sicilian speakers who originated in North Africa and who colonized Malta in both the ninth and mid-eleventh centuries. Though it came under non-Arab rule shortly after this, the Maltese variety of Arabic appears to have been robust and to have resisted much Romance influence until the sixteenth century, whereas Sicilian Arabic came under Romance influence as early as the eleventh century.

4.4.2 Cyprus

Though Arab raids struck the island of Cyprus beginning in the seventh century, it was firmly under Byzantines or European rule for most of its history, and sea raids were largely suppressed by 965 CE. Maronite Christians are said to have moved to Cyprus as early as the eighth century, in the same wave of migration that took many Maronites into Lebanon, but at this point they would probably have still been Aramaic speakers. Following this, the first major migration to Cyprus of

what could have been Arabic speakers was in 938 CE following the destruction of the Maronite monastery in Apamea, some 130km SW of Aleppo, which resulted in the transfer of the Maronite church to Lebanon, but may also have resulted in some migration to Cyprus.

The island was sold to Guy de Lusagne during the crusades at the end of the twelfth century, who invited Maronite settlers to the island, and the final migrations are typically dated to the thirteenth century with the collapse of the crusader states (Borg, *EALL : “Cypriot Maronite Arabic”*, pp. 536-7), but Hourani (1998) suggests that migrations could well have continued after this time, including during times of persecution under the Ottomans starting in the sixteenth century. However, the initial settling populations appear to have moved first from NW Syria and Lebanon to Cyprus between the ninth and thirteenth centuries CE. The linguistic evidence also points to a migrating population whose Arabic was still rather influenced by Aramaic (Borg, *EALL : “Cypriot Maronite Arabic”*, pp. 539-541), suggesting a fairly early date of migration.

4.4.3 Central Asia

The Uzbekistan and Afghan dialects of Arabic appear to date back to quite early Iraqi dialects. Arab armies began the conquest of Central Asia, specifically the areas north and east of the Oxus river (“Transoxania”) in 705 and ended in 750. The armies which conquered and reinforced this area seem to have been largely drawn from Iraq, though they often formed a minority among their various local auxiliary troops (for a blow-by-blow account of the invasion, see Kennedy, 2007, chapter 8) Nonetheless, though there are competing theories for the origins of Arabic speakers in central Asia, the linguistic and historical data supports an eighth century, Iraqi origin for these dialects (Jastrow, 1995; Zimmermann, *EALL :*

“Uzbekistan Arabic”; but see Owens, 2005).

4.4.4 Sudan, Chad, Nigeria

The dialects of the sub-Saharan Arabic dialects in Sudan, Chad and Nigeria appear to be derived primarily from dialects spoken in Upper Egypt, with Arabs firmly established in the region by 1392, though the migrations may have begun earlier. Though a truce had been declared between Nubia and Egypt quite early after the Islamic conquests in the mid-eighth century, various groups that settled near Aswan continued to raid further south.

Invasions from Egypt occupied Nubia first in the late twelfth century, though this was brief, with a more extensive occupation by the Mamluks by 1315, which may have been the primary event allowing Arabs to move south. By 1392, a letter from the ruler of the Kanem empire of Lake Chad complains about the disruptive presence of Arabs to the Mamluk Sultan in Egypt (Owens, 2003, pp. 721-2). In 1504, the Christian kingdom of ‘Alwa, situated just south of where Khartoum now stands, also fell to Arab invaders (Pommerol, *EALL : “Chad Arabic”*).

These dialects, therefore, probably represent the speech of a continuous series of population movements coming from Upper Egypt beginning primarily in the twelfth to sixteenth centuries, though there may have been limited earlier immigration south of Aswan.

4.5 Conclusion

Though it is difficult to summarize this chapter and the preceding one in prose, the map in Figure 4.6 (on the following page) attempts to show all of the early Islamic migrations. The image is a scaleable vector map, which means that

the reader should be able to zoom in as far as necessary in their PDF viewer, without a loss of quality.

Chapter 5

The Linguistic Development and Reconstruction of Arabic Demonstratives

The goal of this chapter is to produce a linguistic analysis of the development of the demonstratives in Arabic, from a historical linguistic perspective grounded in functional syntactic and morphological analysis. The primary focus of this chapter is on the adnominal and pronominal demonstratives in Arabic, as attestations for adverbial demonstratives (whether location or manner adverbials) were inconsistent in the sources. When possible, information about those other demonstrative forms is used to better understand the history of the adnominal/pronominal demonstrative series, but the incomplete nature of the data makes it hard to use them with any consistency. The following chapter will tie the linguistic data to the historical and social data using the theory articulated in Chapter [1](#).

This chapter begins with a typological overview of demonstratives, followed by a look at demonstratives within the Semitic family. Following this, it reconstructs the development of the inherited demonstratives (referred to here as

**ǝa*:-type demonstratives), followed by a reconstruction of the innovative demonstratives which either derived from the original attention-gathering particle **ha*: or are produced by affixing this particle to the inherited **ǝa*:-type demonstratives. The final sections will, as far as is possible, sketch a relative chronology of these changes.

5.1 Typology of Demonstratives

Diessel (1999) is a comprehensive typological study of demonstratives, drawing on a typologically balanced sample of 85 languages. Diessel (*ibid.*, p. 2) defines demonstratives as (spacial) deictic expressions which fulfill specific roles, namely: Syntactic roles as pronouns, nominal modifiers, and adverbs; Pragmatic roles, “focusing the hearer’s attention on objects or locations in the speech situation”, frequently in combination with a physical pointing gesture; Semantic functions, specifically orienting the hearer to the surrounding situation, especially with respect to distance from a deictic center. Diessel (*ibid.*, p. 2) does not require that forms make a contrast of distance in order to be demonstratives, since they can still act to orient a hearer in the surrounding situation.¹

Diessel (1999) distinguishes between demonstratives on two primary levels: First, between different morphosyntactic variants of demonstratives, and then between pragmatic functions.

Morphosyntactically, demonstratives are divided into:

Demonstrative Pronouns Self-standing demonstratives which may be used anaphorically as a pronoun

¹ Dixon (2003) also seems to allow for demonstratives which make no specific distance contrast, though he does not state so explicitly.

Demonstrative Determiners Demonstratives which modify a noun or noun phrase (also called “adnominal demonstratives”)

Demonstrative Adverbs Demonstratives which refer to location or manner

Demonstrative Identifiers Demonstratives which “identify a referent in the speech situation” (Diessel, 1999, 5, see also 5–6, 78–79) and which typically are found in copular clauses

Diessel only uses these terms when referring to categories that are actively distinguished in the language. If he is referring only to the function (e.g. if a language uses the same form for demonstrative pronouns and demonstrative determiners) he refers to these functions as follows: pronomial, adnominal, adverbial and identificational demonstratives. Since the focus of this chapter is not on making a typological comparison, all of those terms will be treated as interchangeable in the discussion of Arabic demonstratives, so for example ‘demonstrative determiner’ and ‘adnominal demonstratives’ will be essentially synonymous.²

Pragmatically, Diessel (*ibid.*) distinguishes demonstrative uses as follows:

Exophoric Referring to entities outside of the discourse

Endophoric Referencing to entities and speech within the world of the discourse

Anaphoric Referring to previously mentioned entities in the discourse

Discourse deictic Referring to propositions or speech acts in the discourse, whether previously mentioned or not-yet-mentioned (cataphoric use)

² I plan to change this in a later draft, but for the time being I will not be making the same distinction as Diessel (*ibid.*)

Recognitional Referring only to unmentioned knowledge shared between interlocutors, typically marked with adnominal demonstratives

Not all languages distinguish all of these categories morphologically, and in most languages in his sample, demonstrative determiners and demonstrative pronouns are either the same or derived from the same bases. Morphologically, pronominal demonstratives are more likely to inflect (for number, gender or case) than adnominal and identificational demonstratives, which are both more likely to inflect than adverbial demonstratives. Pronominal demonstratives usually reflect the same nominal inflection as full nouns, while other demonstratives are less likely to show the same level of marking (Diessel, 1999, pp. 32-33).

Dixon (2003) argues for a separate category, that of VERBAL DEMONSTRATIVES which are essentially the same as Diessel's (1999) "manner adverbial demonstratives," and I see no compelling reason to treat these as a separate class. Manner demonstratives are often used as exophoric demonstratives, e.g. "do it like *this*" with an accompanying hand gesture, or can refer to previous discourse, e.g. "The man gave them an evil stare, the kind that freezes the blood. For hours he looked at them like *that*, before..."

Demonstrative identifiers are a somewhat unusual or restricted class, and are not frequently distinguished formally from other demonstrative types. These demonstratives "focus the hearer's attention on entities in the surrounding situation or in the universe of discourse (Diessel, 1999, p. 79)" and they typically occur in specialized constructions, usually in copular or nonverbal clauses. They must be distinguished, however, from SENTENTIAL DEMONSTRATIVES, a type of presentative, like French *voilà* or Latin *ecce*. Though both forms introduce new discourse topics, demonstrative identifiers are typically more syntactically constrained and

associated with particular constructions, while sentential demonstratives are often “used as one word utterances, which may be loosely adjoined to a neighboring constituent. (Diessel, 1999, p. 79)”. The distinction is not clear cut, and it seems likely that sentential demonstratives may be a source of demonstrative identifiers or visa-versa.

The term demonstrative identifier is not as clear as I would like, so in this chapter, I will refer to presentatives which introduce a single referent into the discourse as a REFERENT PRESENTATIVE, while SENTENTIAL PRESENTATIVE will refer to particles which serve to introduce an entire sentence or proposition and mark it as important. The distinction is similar to the distinction between anaphoric and discourse deictic demonstratives, where one refers to a single referent and the other refers to a proposition, and the terms reflect this distinction transparently. The distinction between these two is also reflected grammatically, since referent presentatives often occupy a slot in the sentence (usually subject), or could be easily construed as occupying such a slot, while sentential presentatives tend to have no syntactic status within a sentence.³ Both types could be described as “presentatives”.

The pragmatic functions described by Diessel (*ibid.*) are generally straightforward, with the exception of discourse deictic and recognitional functions. The discourse deictic function differs from the general anaphoric function in that it refers to propositions within the discourse, as opposed to entities (normally represented by an NP). Discourse deixis is different from true deixis, as in “I’m sorry, I didn’t hear you. Could you repeat *that*?” where *that* here refers specifically to a series of sounds, not a proposition as such, and therefore is a type of exophoric deixis (*ibid.*, p. 101).

³ At least in Arabic, this is not a claim more generally about other languages.

The recognitional use of demonstratives is to activate discourse-new, hearer-old information, that is to bring shared background information into the discourse. The recognitional use is exclusively marked with demonstrative determiners attached to the noun phrase being introduced into the discourse. For example, in “I couldn’t sleep last night. *That* dog (next door) kept me awake.”, the speaker assumes shared knowledge of a noisy neighbor dog, but only activates it in the second sentence. If the listener was not familiar with the speaker’s situation, the sentence would be infelicitous.⁴

Recognitional uses of demonstratives sometimes develop into DETERMINATIVES, a grammaticalization of the recognitional function of a demonstrative into a topic-introducing particle. For example, in the sentence “The true artist is like one of *those* scientists who, from a single bone, can reconstruct an animal’s entire body.” the word *those* refers to something both discourse-new and hearer-new. These often grammaticalize into markers of the nominal heads of relative clauses, since discourse-new/hearer-new information is often further defined in the clause. Determinatives, unlike recognitional demonstratives, may also be used without an accompanying noun phrase, as in “*Those* who backed a similar plan last year hailed the message.” (Diessel, 1999, pp. 108-109, 135-7)

Some languages appear to develop compound demonstratives which fuse presentatives with other demonstrative forms, so Czech has a normal series of demonstratives to which a suffix *hle* can be added to “indicate that one is calling

⁴ If the speaker assumed listener unfamiliarity, they might well use a non-recognitional, topic introducing form which is derived from a demonstrative, *this*, as in, “This dog next door kept me awake.” See Diessel (1999, p. 109) and Ionin (2006) for more on this function of *this* in English and other languages. Note also that *that* is not the only possibly determiner that can be used in this sentence: *The* could be used as well, with a similar meaning (see Gundel, Hedberg, and Zacharski, 1993).

special attention to the object being pointed out (Imai, 2003, pp. 65-6).” This particle appears to be derived from an earlier presentative that can be translated as “behold”.

There is some contention as to whether demonstrative systems must have more than one term which is marked for distance, though both Diessel (1997) and Dixon (2003) appear to allow for such a case in their definitions. Anderson and Keenan (1985, p. 280) state that while, in principle, a language could have a single demonstrative that means something like “present to speaker” or “present in the extralinguistic context of the utterance,” such a system would be “little different from a definite article.” Nonetheless, in some languages, such as Czech and (Cairene) Egyptian Arabic only one demonstrative is used with any regularity.

Furthermore, if one looks at the pragmatic functions performed by demonstratives as described above, not all of these actually require multiple terms marked for distance — anaphoric uses (referring to entities present in the previous discourse) and recognitional uses do not require two terms, since in the case of anaphoric uses there is no need to point in multiple directions, as with discourse deixis, and in recognitional uses there is no need to indicate distance when referring to shared knowledge.

Hence, a single term demonstrative system could effectively fulfill many of the functions of demonstratives. Such demonstratives would indeed differ significantly from a definite article, contra Anderson and Keenan (*ibid.*, p. 280) — they might resemble pronouns in their anaphoric use (for single term demonstratives with pronomial e.g. anaphoric use in Semitic, see Huehnergard and Pat-El, 2012), or be used to activate shared knowledge as in referential use, like the definite article, though that typically indexes hearer-old, discourse-old information, as opposed to referential use which indexes hearer-old, discourse new information.

Similarly, a presentative need not distinguish distance per se (more so for offeratives, which rarely distinguish direction), which could be marked with additional locative demonstratives.

Single distance demonstrative systems can even be used for discourse anaphora: Dixon (2003, p. 83) notes that Supryire (Niger-Congo) uses demonstratives from anaphoric reference, and a demonstrative with a third person pronoun for cataphoric reference, even though the demonstrative system only marks one level of distance.

In terms of their grammaticalization, demonstratives appear to follow a different path than most function words. Diessel (2012) argues against a view in grammaticalization theory that all functional words have ultimately lexical origins, and suggests instead that demonstratives instead are “a unique class of linguistic expressions providing another frequent source for the development of grammatical markers (p. 38).” He refers instead to Bühler’s (1934) two-field theory of demonstratives, which distinguishes between “deictic” functions and “naming” functions of words. Most words in a language fulfill naming functions, while, true deictic words fulfill the (exophoric) pointing function. These deictic words are original “indeclinable words with no specific semantic and/or syntactic features and functions (ibid., p. 41)” which only later develop into things like demonstrative pronouns, which fulfill both a deictic and a naming function, and which Bühler (1934) refers to as mixed forms.

These mixed forms do not entirely lack their original pointing, deitic nature, as made clear in Bühler’s (1934) account of this process:

The ‘pure’ deictic signal is and was, when it occurs and occurred, or would be if it occurred, an arrow showing the way without a name

written on it, and nothing more; the arrow function on the signpost is not cancelled by painting a place-name on it; nor is it cancelled when the German word *dér* (this or the) emerged from the particles of the *to-Deixis: The word ‘der’ has at least so much in the way of the naming function entrusted to it that it can find a place in the symbolic field among the other naming words; hence its correct name, pronoun. (quoted in Diessel, 2012, p. 42)

Later processes act to metaphorically extend the use of this pointing function, resulting in the various types of endophoric functions explored by Diessel (1999). They may further grammaticalize, losing any pointing function at all, and become little more than syntactic markers, as in the case of determinatives.

Finally, while demonstratives are typically treated in terms of distance marked, with almost all languages regularly marking demonstratives for something akin to ‘near to speaker’ and ‘far from speaker,’ actual distance is rarely the determining factor in the choice of demonstrative forms. Instead, it typically is based on degree of speaker control over the object (Imai, 2003; for examples in Arabic, see Jarbou, 2010). For endophoric use, the choice of adnominal demonstratives tends to be determined by the referential status of the modified known in terms of the givenness hierarchy, as discussed in Gundel, Bassene, et al. (2010); Gundel, Hedberg, and Zacharski (1993). Nonetheless, for the purposes of description here I will retain the traditional notion of ‘proximal’ and “distal” demonstratives as they conveniently subsume a more complex range of phenomena.

5.1.1 Arabic Demonstratives from a Typological Perspective

This section gives an overview of demonstratives common within attested Arabic dialects, including Classical and Quranic Arabic, within the typological

framework established in the previous section. Much of the general typology of Arabic demonstratives applies also to Semitic languages generally, explored below in greater detail, but all examples will be given using Arabic forms from my sample.

Arabic demonstratives typically only distinguish two degrees of distance contrast⁵, and in almost all dialects demonstrative pronouns are inflected for gender and number.⁶ Most dialects make no gender contrast in the plural, but some do make such a distinction, and it is typically true that a dialect that makes a distinction between masculine and feminine plural in the demonstratives makes the same distinction in the pronouns, though the opposite is not always true.

Table 5.1: Pronominal and adnominal demonstrative forms from CHDSH.

	m. sg.	f. sg.	m. pl	f. pl
Proximal	da:	di:	do:l	de:l
Distal	da:k	di:k	do:la(:)k	de:la(:)k

In many dialects, the form used for demonstrative pronouns and demonstrative determiners are identical, however, some dialects have two series of demonstrative determiners — one series is typically identical to the demonstrative pronouns, while the other series is either derived from the demonstrative pronouns or another source, and used exclusively as demonstrative determiners. In Moroccan Arabic, for example, as will be discussed at greater length below, the determiner

⁵ There are a small number of dialects which appear to have innovated a medial degree contrast, the only one in my sample being ANDAR.

⁶ Arabic distinguishes two genders, masculine and feminine. In modern dialects, agreement marking (on verbs, adjectives, demonstratives) only distinguishes between singular and plural, but nouns may be marked with an ending that indicates duality, though co-referent verbs and nouns would typically take plural agreement. Classical and Quranic Arabic did mark dual agreement on verbs, adjectives, and demonstratives.

series is a grammaticalized version of the demonstrative pronouns which have lost their gender and number agreement in the proximal series due to being prefixed to nominals, since gender was marked word finally. The distal demonstrative determiners, however, still show agreement since the gender and number information was carried word internally.⁷ However, full versions of the demonstratives can still be used adnominally, often with a more specific or marked meaning (explored in much greater detail in Brustad, 2000, and below), as in the following examples:.

- (1) (Moroccan Arabic, from Caubet, 1993, p. 168)

fnu: ha:d-əf-fi
what this-DEF-thing

‘What’s this (thing)?’

- (2) (Moroccan Arabic, from Brustad, 2000, p. 130)

w fħal d l-ħba:b yʃarfu:ni ʃla f-fi: ha:d-a
and how.many of DEF-friends know for DEF-thing this-M.SG

‘And so many friends know me for this (particular) thing’⁸

Dialects vary in whether they prepose (Dem-N) or postpose (N-Dem) the demonstrative determiners, and typically there is some fluidity, so that a dialect that postposes the demonstrative determiner can prepose it for rhetorical effect,

⁷ The vowels that mark gender and number are from the same source, but the distal series had a final consonant *-k* (marking the series as distal) which protected the vowels.

⁸ It is particularly interesting in that *ha:d-əf-fi:* is typically used as a grammaticalized inanimate demonstrative pronoun, while only animate referents are substituted by a gender and number marked demonstrative pronoun (see Caubet, 1993, p. 168), so the inversion of this phrase in the second example shows that this grammaticalization process is still ongoing.

as in Cairene Arabic, where in the second example the preposed demonstrative emphasizes the unpleasantness of the ‘question’:⁹

- (3) (Cairene Arabic, from Doss, 1979, p. 350)

fof-t er-ra:gel da:?
saw-you DEF-man this.M.SG

‘Did you see this man?’

- (4) (Cairene Arabic, from *ibid.*, p. 350)

seb-na min di: s-si:ra
leave.IMP-us from this.F.SG DEF-question

‘Let us get over this question!’

Arabic dialects typically distinguishes near and far locational demonstrative adverbs and generally have a manner adverb. Far demonstratives are often, but not always, derived transparently from the near demonstrative, so that the southern Syrian dialect LvS433 has the following near, far and manner adverbs respectively: *ho:n*, *hna:k*, *he:k* while the eastern Syrian dialect of Palmyra LvSPAL has *ho:n*, *ya:d(i)*, *he:k* where the far demonstrative adverb is not derived from the near.

Though masculine forms are often morphologically less complex than feminine forms in Arabic,¹⁰ there is also a tendency to generalize feminine forms into

⁹ This simply seems to be related to markedness and frequency of use, so that whatever the typical order is in the dialect, the atypical order carries very similar meanings to the atypical order in other dialects. That is to say, this isn’t an extraction or fronting process, but rather it seems to be a result of the implicature generated by using the less frequent structure.

¹⁰ For example, almost all nouns which are not marked with a final *-a(t)* morpheme are masculine (for exceptions see S. Procházka, 2004), and the past tense masculine verb forms are the stem form from which other past tense forms are derived.

gender neutral forms. This is partially a result of a complex interaction between animacy and gender — poorly individuated groups of inanimate referents are often treated grammatically as a feminine singular entity, and even groups of animate referents treated as a collective may be treated as feminine singular in the grammar (this topic is treated extensively in Brustad, 2000). The same is true for demonstratives, so e.g. Cowell (1964, pp. 553-4) reports for Levantine dialects that:

Since masculine (/singular) is the neutral or bass [sic] number/gender, the masculine demonstrative is generally used in reference to an object whose name the speaker does not know and commonly also as sequent to a clausal antecedent [i.e. discourse deitic — AM] or a vague or conceptual antecedent. The feminine, however, is also commonly used in the latter case: *fu: hayye?* ([what this.fem]) ‘What’s this, i.e. What’s up? What’s happening?’, *hayy hiyye* ([this.fem she]) ‘That’s it! i.e. You’ve hit the nail on the head!’

Thus, it might be said that when paradigmatic levelling occurs in Arabic, both the masculine and feminine forms are approximately equally likely candidates to be levelled across the paradigm, especially when plurality is involved.

In Classical Arabic, which has a case-system distinguishing between nominative, genitive and accusative cases with a set of suffixes (some of which are synthetic), demonstratives generally do not show case marking in Classical Arabic. The exception to this are dual demonstratives, which had the forms shown in Table 5.2

Table 5.2: Classical Arabic dual demonstratives

	Proximal		Distal	
	NOM	OBL.	NOM	OBL
MASC.DUAL	ha:ða:ni	ha:ðayni	ða:nnika	ðaynnika
FEM.DUAL	ha:ta:ni	ha:tayni	ta:nnika	taynnika

5.2 Demonstratives in Semitic

The most complete and recent reconstruction of Semitic demonstratives is Hasselbach (2007a). She reconstructs the demonstratives as being composed of base demonstratives augmented with a series of affixes which are combined in various ways in different languages, at times in different orders.

The primary marker of demonstratives, common to most of the Semitic languages, is an element **ð* in the singular, with the plural element *ʔul(l)i*.¹¹ These two elements perform the primary function of the near demonstrative (pronominal and adnominal), obviously inflected for number. Inflection for gender in the singular originally appear to have been via the feminine ending *-a:t*, though by West Semitic Hasselbach (*ibid.*, p. 22) argues the distinction was simply in the vowel alternation of masculine *ði:* vs. feminine *ða:*. The proto-Semitic and proto-West Semitic demonstratives are only proximal, with demonstrative forms that later developed into the third person pronouns frequently taking the role of far demonstratives (see Huehnergard and Pat-El, 2012). The base near demonstratives in proto-Semitic then are as follows:

¹¹ The vowel quality in the plural element is not entirely clear from the Semitic evidence, though Hasselbach (*ibid.*, p. 23) argues for an original /u/. Given the evidence of Arabic dialects discussed in Section 5.3.2.1, however, it is not clear if there was necessarily a single form.

Table 5.3: Common Semitic (Proximal) Demonstratives (based on Hasselbach, 2007b)

	m s	f s	m d	f d	c pl
nom.	*ðu:	*ða:tu	*ða:	*ða:tu	
gen.	*ði:	*ða:ti	*ðay	*ðayti	*ʔul
acc.	*ða:	*ða:ta			

The demonstratives in proto-Semitic were inflected for case, and for dual number, but by West Semitic they likely lost most of this inflection:

Table 5.4: Common West Semitic (Proximal) Demonstratives (based on Hasselbach, 2007a)

m s	f s	c pl
ði:	ða:	ʔul(li)

The suffix *-k(a)* is one of the most widespread and productive affixes for the demonstratives, and primarily acts to create far demonstratives from near demonstratives. Its usage does not coincide cleanly with any of the major subgroupings of Semitic languages, but rather is used in some but not all of the Semitic languages, suggesting a proto-Semitic origin. Hasselbach (2007a, pp. 15-16), divides the Semitic languages into three types. The first exclusively use anaphoric pronouns for distal demonstratives, and includes Tigrinya (an Ethiopic language), OSA, Hebrew and Phoenician. The second group consists of languages which have far demonstratives derived from near demonstratives, including Arabic, some varieties of Aramaic, and MSA. The third group, including Akkadian and Ge‘ez have both anaphoric pronouns and far demonstratives based on near demonstratives, but the demonstrative forms are only allowed in restricted contexts, suggesting that they represent an older layer, while the use of anaphoric pronouns as demon-

stratives is probably an innovation in those languages (but see Huehnergard and Pat-El, 2012)

Arabic almost exclusively uses the *k(a)* suffix to generate distal demonstratives from proximal demonstratives. In some varieties (esp. Classical Arabic and the Jabal Rāziḥ dialect, described in Watson et al., 2006a), the *k(a)* suffix appears to be reinterpreted as reflecting the second person pronominal suffixes, so that the demonstratives can ‘agree’ with the gender and number of the addressee. This is found in other languages (see Dixon, 2003, p. 65), but within Semitic appears to be restricted to Arabic and therefore cannot be reconstructed for the other Semitic languages (Rebecca Hasselbach, p.c.)

Hasselbach (2007a, pp. 21-22) argues that the *-k* suffix largely replaced an older suffix *-l(V)* which is attested in Afroasiatic and in some of the Semitic languages, especially in Arabic and its dialects and which she claims marked distal or remote distance. This suffix is found in the Classical Arabic distal singular demonstratives (*ḍalika*, *tilka*), but is absent from almost every modern dialect. In Hebrew and Aramaic, a suffix *l(V)* augments the plural forms, so in the Pentateuch, the plural near demonstrative is *ʔe:lle* but a form *ha:-ʔe:l* (DEF-this.pl) is also found (ibid., fn. 58). In Biblical Aramaic, Hasselbach (ibid., p. 15) gives both *ʔe:lle(h)* and *ʔe:l* for the near plural demonstrative, but I am unable to determine the distribution of these two forms. In Afro-Asiatic, we have a demonstrative base **ll* connected to remote deixis in the Cushitic language Rendille. In Hebrew, we also have a form which becomes more prevalent in later varieties for use as singular near demonstratives, *hal-l-a:ze/e:zû/a:z* (DEF-la-this.m/f/c.sg), but it is not clear if this should be connected with the *l(V)* forms used in the plural suffixes.¹²

¹² Huehnergard (1983) connects the *la* in the *halla:z* forms with the ‘asseverative *lām*’ though as discussed elsewhere, he doesn’t make a strong argument for this position. See also Testen (1998).

There are several issues with Hasselbach's analysis of the *l(V)* as a distal marker: Remote deixis is uncommon in Semitic languages, primarily attested in Amharic. This is probably an areal feature that led to Amharic developing a remote deixis independently: the WALS typological atlas, feature 41a "Distance Contrasts in Demonstratives" shows that East African languages generally has 3 or 4-way distance contrasts, so this might be an areal feature, though the density of sampling is too low for this to be very clear. The accompanying article notes that Africa has an average of 2.45 levels of deictic distinctions, while Asia and Europe average 2.17 and 2.12 respectively. Within Arabic dialects, there are a small number which have developed remote deixis, especially those in north-west Mesopotamia (Jastrow, 1978), but they use a suffix *-a:* that has no *l* component.¹³

The primary piece of evidence that (Hasselbach, 2007a, fn. 41) presents for an analysis of *-l(V)* as a marker of remote deixis is from the Arabic grammatical tradition by way of Wright (1896-1898, p. i. 267):

Some grammarians assert that there is a slight difference of meaning between *dāka* and *dālīka*, the former referring in their opinion to the nearer of two distant objects, the latter to the more remote.

Jarbou (2012) provides extensive arguments against this view. He shows that grammarians were essentially split between two analyses for the variation between dialects that had or did not have *-l-* in their distal singular demonstratives.

¹³ What if any impetus there was for this change is unclear. I have not been able to find evidence for three-way deixis in either Kurdish or other western Iranian languages — three-way deixis is actually seen as an isogloss separating Eastern Iranian languages (much further east than Mesopotamia) from Western Iranian languages which have a binary contrast.

Early grammarians attributed the difference to dialect variation¹⁴, while only in the twelfth century do the grammarians begin to suggest instead that it represents some kind of middle deixis. It would seem, then, that the later Arabic grammarians, in attempting to understand the dialect variation that had been codified into Classical Arabic, created an ad hoc explanation which is still being repeated today.¹⁵ Moreover, the presence of these forms in Classical Arabic distal demonstratives is probably coincidental — many dialects show a layering as a result of contact with other dialects (see Section 5.3.1.1 and Chapter 6), and the singular demonstrative forms with *l(V)* are preserved as the distal series by coincidence, not because the *l(V)* is necessarily distal.¹⁶

Huehnergard (1983) suggests that the *-l-* found in singular demonstratives in Classical Arabic and in Hebrew is actually a reflection of the asseverative *la-*,¹⁷ though his does not provide a strong argument for this assertion.¹⁸ He gives examples of doublets in Classical Arabic such as *ḏa:lika* vs. *ḏa:ka* where the former is more specific in his gloss ‘that very one’ versus that latter ‘that one,’ with a similar pair in Hebrew *halla:z(eh)* vs. *haz-zeh*. Though he does not provide any evidence that one usage is actually more emphatic than the other, this is an intriguing possibility. The fact that the Quran is the only place where we see demonstrative forms

¹⁴ Between the supposedly “Hijāzi” vs. “Tamīmi” dialects, a linguistic-historiographical attribution which needs to be better explored, especially given the much greater and more prominent variation already located in Yemen and the Hijaz.

¹⁵ For a parallel in Modern Standard Arabic, see Michalski (2011) who shows that Classical adnominal demonstratives which agreed with the addressee have been reinterpreted as agreeing with the nominal instead.

¹⁶ See also Hasselbach’s (2007) chart of Classical Arabic distal forms without *-l(V)-* (Hasselbach, 2007a, p. 9).

¹⁷ That is, a particle whose function is to assert the value of the following clause, found in Arabic in examples such as *la-’antum ’ašaddu rahbatan* ‘la-you.pl more fearsome, i.e. you are indeed most feared’ from Wright (1896-1898, pp. I, 283).

¹⁸ Though Testen (1998) treats this particle extensively, he doesn’t go into any detail about what function this *-l-* would have within a demonstrative.

with *-l-* in the singular may support the idea also — the highly emphatic nature of the Quranic text and of the poetic register might encourage the use of these forms. Otherwise, demonstrative forms which contain an *-l-* element in the singular are not attested either in my data, nor in any of the dialect atlases, including Yemen, which has the greatest diversity of demonstrative forms, nor in W. Fischer (1959).

Watson et al. (2006b) does find, in a NW Yemeni mountain dialect, an emphatic series of demonstratives with suffixed *-al* which may be related to the asseverative usage of pre-clausal *la* from Huehnergard's study. These demonstratives do contrast between a general and a specific connotation, in addition to having a frequent presentative use, so that they contrast the sentence *bu: ta:k is-sanat* (in that DEF-year) 'in that year' with *bu: ta:k-as sanah* (in that-/al/ year.INDEF) 'in the same year', much as suggested by Huehnergard for the distinction in Classical Arabic.

In these demonstratives, we can also posit a simple explanation for the shift from what was generally a clause-, or at least word-initial particle to a suffix form: they may have been bridged by a form *w-al* which also functions as a sentential presentative, and which is clearly formed from the word for 'and' combined with this asseverative particle (presumably from the following clause). Speakers would then have reanalyzed this form from being sentence initial (coming after the conjunction) to being a suffix and transferred it to the demonstratives. Alternatively, it might have come after presentative demonstratives meaning e.g. 'look, this indeed is + S' and cliticized. In any case, this *-l* still comes outside of the core of the demonstrative, that is it does not intrude between the **ǝV* base and the *-k* suffix as it does in Classical Arabic. Moreover, this seems to be a phenomenon that is restricted to this unusual dialect which shows so many features peripheral to Arabic dialects that in another article Watson et al. (2006a) question whether it is an

Arabic dialect at all.

To sum up with regard to the $-l(V)$ forms, the function of these forms is extremely unclear — they clearly do not mark distant and definitely not remote deixis. They may be related to asservative uses of $l(V)$, but how this came to be grammaticalized to the end of a demonstrative is not clear.

The suffix $-n$ is similar to $-k$ in that it attaches to demonstratives in a scattered group of languages, but its function is less clear. It is associated with near demonstratives, particularly near masculine singular demonstratives. In Phoenician, for example, some inscriptions contrast zn with z with masculine nouns, and this may be marking a difference in distance, with zn marking closer distances against z , though the normal far demonstrative is the third person pronoun $h?$ (Hasselbach, 2007a, p. 13). Ge'ez, OSA, and Aramaic only have the n in the masculine, while MSA, ANA and Maltese (based on Hasselbach's data) have n in both the m.s. and f.s. near pronouns. In any case, it is likely that the original placement of n was on near masculine demonstratives, though the function is not clear.

Greenberg (1960) sees this use of $-n$ as part of a general Afro-asiatic pattern of marking masculine with a final $-n$, feminine with $-t$ and plural with an unrelated $-n$ suffix (which he uses to explain the presence of $-n$ in plural demonstratives in Semitic, e.g. OSA iln). Though he shows quite a bit of evidence in support of this view, he does not make a clear claim of what kind of part of speech these markers are and so it's hard to understand their exact grammatical development. Nor does his explanation account for the absence of these forms in the distal demonstratives.

It is not clear whether this is the same n which is attached directly to ha in the Hebrew definite article han , and in the Akkadian demonstratives, though this seems plausible. There is also a variant n , $-na$: which is attested primarily in Aramaic, and possibly in Maltese. If that form with a long vowel is restricted to

Aramaic, then it could be an innovation of Aramaic, while if it is found in Arabic it would probably need to be reconstructed as far back as Central Semitic.

A *ha(:)-* prefix exists in a number of Semitic languages, and seems to have originally had some sort of presentative or interjective function (Hasselbach, 2007a, p. 2; Pat-El, 2009, pp. 40-41). This *ha(:)-* acts as (a usually obligatory) prefix on a demonstrative base, usually derived from δV , in a number of Semitic languages, including Akkadian, Ugaritic, Biblical Hebrew, and most varieties of Aramaic after the Middle Aramaic period. It is not used in many Modern South Arabian and Ethiopic Semitic languages, and does not frequently appear on the demonstratives in Old South Arabian and Phoenician (though see Pat-El, 2009, pp. 45-6). The use of this prefix is sporadic in Middle Aramaic dialects (more details in *ibid.*, p. 46), but the distribution tends to be that demonstrative pronouns do *not* have *ha(:)-*, while demonstratives which can either act as determiners or pronouns do have a *ha(:)-* prefix (Hasselbach, 2007a, p. 14; Pat-El, 2009, pp. 22-3). For this reason, Hasselbach (2007a, p. 20) argues that the “demonstrative particle **ha(:)-*, including its extended form **han-* was originally used to mark adnominal function [in demonstratives].”

The Common Semitic demonstratives did not mark gender in the plural, and thus each language that has gender-differentiated plural demonstratives likely innovated these individually, often in analogy with the pronouns (*ibid.*, p. 16). The form of the proto-Semitic plural demonstrative is $?VI(lV)-$ where the vowel is either a high front vowel /i/ (Hebrew, Aramaic) or a high back vowel /u/ (Arabic, Akkadian), and several languages are ambiguous.¹⁹

¹⁹ See Section 5.3.2.1 below for a discussion of this vowel.

Table 5.5: Summary of Semitic Demonstrative Formatives

Formative	Function	Languages with that form
<i>-k</i>	Distal marker	Ge'ez, MSA, Arabic, Ugaritic, Aramaic
<i>ha:-</i>	Intensifier, Marking adnominal demonstratives	Akkadian, Ugaritic, (later) Hebrew, Middle and later Aramaic
<i>-l(V)</i>	Unclear	Classical Arabic (in singular distal), Jabal Rāziḥ Arabic, Hebrew (singular proximal), Aramaic (plural)
<i>-n</i>	Masculine marker (proximal)	Ge'ez, MSA (Ḥarsūsi), Maltese(?), OSA, Aramaic, Phoenician

5.2.1 Functions of Semitic Demonstratives

These reconstructed demonstrative forms in Semitic fulfill many of the expected functions of demonstratives: They operate exophorically and endophorically, as anaphors (or as discourse deitics). As pronouns, they can of course act as the subject in a copular sentences, though Semitic languages generally do not distinguish a class of demonstrative identifiers morphologically distinct from their demonstrative pronouns.

It is also necessary to reconstruct back to proto-Semitic forms derived from the demonstrative which act as the first term in either a possessive phrase or even in a type of relative construction (Huehnergard, 2006). These are typically referred to in the Semitics literature as DETERMINATIVE-RELATIVE PRONOUNS (det.-rel. pro-

nouns), and typically translate as something like ‘the one of’ or ‘the one who’. They can be preceded by a nominal, in which case they presumably operate as a kind of apposition to that nominal, i.e. ‘N, that one of.’ Huehnergard (2006, p. 114) gives the following examples for proto-West Semitic forms (in the nominative):

(5) Without an antecedent

- a. *ðu:* *bayt-i-m*
 one.NOM house-GEN-UNBOUND
 The-one-of the house
- b. *ðu:* *naðara*
 one.NOM guarded.3ms
 The-one-who guarded

(6) With an antecedent

- a. *baʕl-u-m* *ðu:* *bayt-i-m*
 lord-NOM-UNBOUND one.NOM house-GEN-UNBOUND
 The lord (the-one-) of the house
- b. *baʕl-u-m* *ðu:* *naðara*
 lord-NOM-UNBOUND one.NOM guarded.3ms
 The lord (the-one-) who guarded

In many of these languages, these det.-rel. pronouns are similar in form to the demonstratives, though they inflect for case, which demonstratives generally do not do. There is often a single form leveled for all genders and numbers, e.g. Aramaic *ði:* and Ethiopic *za-*. At the same time, developments within the det.-rel. paradigm seem to have occurred independently of the developments in demonstratives, so that in Classical Arabic the plural forms *ðawu:* and *ðawa:tu* are clearly formed by augmenting the base forms with endings based on the sound plurals endings (*ibid.*, fn. 59)

For some of the Semitic languages, these forms also act as relative clause markers, though others have innovated new sets of relative markers, as is true for most, but not all Arabic dialects. Thus, the Semitic languages could be said to lump genitive and relative clauses, a feature which is not uncommon worldwide (see Hendery, 2012, §2.2.4)²⁰

5.2.2 Semitic and Arabic demonstrative development

For many of the developments which will be described in the following sections, there are numerous parallels among the Semitic languages. At the same time, however, many of these developments are widely separated in time and space, and suggest a type of parallel evolution that is not necessarily related to descent from a common ancestor. Instead, given that many of the morphological forms which we will describe actually arise through syntax (i.e. via grammaticalization), what is inherited in many of these languages is a related, common set of frequent syntactic constructions, along with various free and bound morphemes which operate in similar ways and which therefore follow similar grammaticalization pathways.²¹ As Joseph (2012, p. 160) eloquently states:

In particular, if one looks to variation in the proto-language for the languages involved as a source of parallel developments [...] then parallel developments can reflect inheritances into each language of variants that existed in the proto-language. Thus one has cognacy, to be sure, either in actual forms that are used grammatically or at the more ab-

²⁰ The WALS database, however, for the feature 60A *The differentiation of genitive, relative and adjectival phrases*, makes this lumping appear relatively rare, showing up only in Neo-Aramaic and Burmese out of , out of 138 languages surveyed.

²¹ One can even extend this notion of similarity of structure to understanding convergent evolution between unrelated languages, i.e. Enfield's (2003) "typological poise".

stract level of patterns or categories with different variants as their realization. However, the variants, if they make their way into each language, can give the impression of copies. Moreover, the timing for the “activation” that is to say the emergence, of a variant within a language is independent of its activation/emergence in another, so that one sees what amount to parallel but independent developments that are nonetheless inherited.

One of the clearest examples of this is the development of the demonstratives which prefix a particle **ha:/*han* to demonstrative forms derived from δV . This structure is attested in the Akkadian, that is as early as the third millennium BCE, and in Biblical Hebrew, dating from the first millennium BCE. The attestation of demonstratives with this particle are scattered within Aramaic. Though the Aramaic group is attested from nearly the 9th century BCE, **ha:-*initial demonstratives appearing starting from the Jewish Palestinian Aramaic dialects around the third century CE.²²

The development of **ha:-*ða:* demonstratives in Arabic seems to be more recent than in many of those languages. The dialects in North Africa, which arrived there at earliest in the seventh century CE, have two sets of demonstratives — demonstrative pronouns, and demonstrative determiners, which are prefixed to a following noun. In many North African dialects (e.g. NADJ whose forms will be shown here), only the proximal demonstrative determiner (*ha:d-*) has the **ha:-* prefix. The distal demonstrative determiners *da:k/di:k/du:k* have no such prefix. However, both the proximal and the distal demonstrative pronouns have the **ha:-* prefix, so one finds proximal *ha:da* and distal *ha:da:k*.

²² This dating is largely based on the paradigms presented in Hasselbach (2007a).

The difference between the demonstrative pronouns and the demonstrative determiners in this dialect suggest that two processes were happening in the ancestors of North African dialects: the demonstratives which had previously been used both for demonstrative pronouns and determiners were slowly developing determiner forms as a result of grammaticalization. At the same time, the **ha:-* prefix was only just beginning to diffuse throughout the demonstrative paradigm. The determiners seem to have finished forming before the **ha:-* prefix had fully diffused, and so there is no **ha:-* prefix in the determiners.

Since both these developments are essentially restricted to North African dialects, it suggests that they are quite recent, probably starting shortly before the seventh century CE in the predecessors to North African dialects. Thus, the development of **ha:-*prefixed demonstratives seems to have begun only around the seventh century in those dialects, and in many dialects it never occurred at all, as in the dialects of the Nile Valley. The development of the **ha:-*prefixed demonstratives is therefore much later than in other Semitic languages.

There is also little evidence that **ha:-*initial demonstrative forms were inherited but lost, that is that **ha:-*attachment actually was an old development that is disappearing. Few dialects seem to show the kind of phonological erosion that would suggest a gradual loss of **ha:-*, and where there *are* dialects like this, they are generally surrounded by other **ha:-*demonstrative dialects, and are often in contact with a foreign language. Dialects with clearly eroding **ha:-*prefixes from my sample are ANDAR, an Anatolian Arabic dialect, CYP²³ and the North African NATJ. Indeed, the data and dialect geography all seem to indicate that the process of **ha:-*prefixation is increasing the number of dialects with **ha:-*prefixed dialects,

²³ Indeed, the loss of the /h/ in Cypriot demonstratives is unusual, since /h/ normally merges with /x/, not with /Ø/, i.e. it isn't normally lost.

rather than the opposite, which is what we would expect from a scenario of inheritance and loss.

Similarly, the process of attachment of these forms also seems to differ. In Biblical Hebrew and Aramaic, the process of the **ha(n)* prefixes was tied closely to the process of the development of the definite article, as describe in Pat-El (2009). These prefixes in that analysis would have marked attributive elements (including demonstratives) in a sentence, and from there were reinterpreted as a definite article. This is reflected in the distribution of **ha(n)*-initial demonstratives in Jewish Palestinian Aramaic, where pronominal demonstratives have no prefix, but adnominals must have it.

In Arabic, however, the opposite situation is found, as discussed previously with regard to North African dialects — in the distal series, **ha:-*initial demonstratives are pronominal, whereas the adnominal series have no prefix. It would seem, then, that the **ha:-*prefix which is attached to Arabic demonstratives is not a marker of attributive status, but performed a different function, as will be discussed at length below.

Moreover, the prefixing of **ha:-* to Arabic demonstratives does not seem to have happened at the same time as the development of the definite article, as it was in Biblical Hebrew and in Aramaic. The definite article is widespread throughout Arabic dialects, and there are very few Arabic dialects (if any) whose definite article isn't derived from apparently the same source which produced the widespread definite article **ʔal-*.²⁴ However, there is great diversity in the attachment of **ha:-* to demonstratives, suggesting that the diffusion of the definite article is a much

²⁴ The relatively rare variants of the definite article, *an-* and *am-* are probably ultimately derived from the same source as *ʔal-*.

earlier process than the rather late process that led to **ha:- *ǝa:* type demonstratives in Arabic.

These arguments all suggest that the development of the demonstratives in Arabic, while parallel and similar to the developments in other Semitic languages, are not the result of a common descent, or even of identical processes of development. Instead, each language underwent its own path of change in the demonstratives, on its own timeline, using similar material (e.g. **ǝa:*-type demonstratives with their many functions, and the similarly versatile **ha:* prefixes) to produce similar results.

For this reason, the focus on this chapter will be specifically on the developments that occurred in the ancestors of the modern Arabic dialects. The processes may or may not have been the same in other Semitic languages, but as the evidence shows Arabic likely underwent many of these changes independently of the other languages in its family, at different times. Where a contact explanation is called for, or where there is an especially clear and helpful parallel to another Semitic language, the analysis will make note of this, but the reconstruction of the paths of change will be specific to the Arabic language.

5.3 Demonstratives derived from **ǝa:*

The ancestors of all modern Arabic dialects featured demonstrative forms derived from an original Semitic set of demonstratives that had forms derived from original singular **ǝV* and plural **Vl(l)V*. However, there have been a number of innovative changes in Arabic dialects that have transformed these demonstratives. These innovations are not entirely morphological (that is to say, motivated by paradigmatic pressures) but are also the result of grammaticalization

processes, which requires understanding the functional syntactic roles played by the elements which have now become parts of the demonstratives.

5.3.1 Development of singulars

5.3.1.1 Gender differentiation

There is a good amount of evidence to suggest that, although the PWS demonstratives **ǝa:* (feminine in PWS) and **ǝi:* (masculine) are marked for gender, this was generalized in an archaic variety of Arabic (what could be called proto-Arabic) to *ǝa:* (see also W. Fischer, 1959, p. 35). Given the shared history of Araboid languages, this is a plausible scenario for the development of the older Arabic demonstratives. Of course, it is still necessary to explain Such a linguistic event would explain two aspects of the Arabic data, the ‘flip’ of gender marking vowels in Arabic vs. PWS, and second the existence of dialects with no vowel alternation. This change has happened more than once within the history of Arabic dialects, so that e.g. LvSPAL and LvSSOU both have demonstratives which are not distinguished for gender in the singular (as a result of relatively recent sound change), and North African dialects have a gender-less proximal demonstrative determiners, in addition to the various dialects which still attest remnants of a gender-less singular demonstrative.²⁵

One reader also objects that the reconstruction of a gender-less demonstratives would have left this early stage of Arabic with gender marking in the relative, but not in the demonstratives. However, there is no clear reason to reconstruct *?allaǝi:/ti:* (REL.M/REL.F) to the ancestor of modern dialects, since the vast majority

²⁵ It is also interesting to note that both LvSPAL and LvSSOU have gender-marked pronouns in both the singular and the plural, but have not restored gender marking to their demonstratives following that sound change which merged the masculine and the feminine.

of these dialects either have forms which do not mark gender (e.g. *ʔilli*) or which are apparently derived from gendered pronouns directly (e.g. the dialect of Abha which has (rarely) m. *ǰa:*, f. *ta:* and c.pl. *ʔilli*, though the plural form is used in most cases.) Finally, if the analysis in this chapter is correct, the *ʔalla-ǰi:-ti:* forms could only have developed *after* the leveling of *ǰa:* as a common singular form. As discussed previously, feminine forms are just as likely to be levelled as the masculine forms, so a scenario in which levelling of original feminine **ǰa:* occurred is not surprising.

Dadanitic (ANA) data suggests another possible scenario for the development of a general *ǰa:* demonstrative. In Dadanitic, the masculine demonstrative adjective is *ḏḥ*, probably *ǰa:*, while the feminine is *ḏṭ*, likely *ǰa:t* (Macdonald, 2008a, p. 209). A sound change that occurs in Arabic, where feminine *-t* is lost except in construct (i.e. lost ‘in pause’ or phrase finally) would have caused a merger of these two demonstratives to *ǰa:*. Remnants of the form with original *t* are found in Classical Arabic’s det.-rel. feminine singular pronoun *ǰa:t*, which is always in construct by virtue of its function. The CA det.-rel. certainly seems to preserve an early form, since we would expect to find in a vowel-alternating dialect something based on *ǰi:* for the feminine, such as *ǰi:yat* or *ǰi:ʔat*, or in a consonant-alternating dialect *ta:t*.²⁶ Given the shared history of Araboid languages, this is also a plausible scenario.

Whatever its origin, the demonstrative **ǰa:* would have been undifferentiated for gender in an early stage of Arabic, but later on, either in analogy with pronouns and the rest of the gender system, dialects innovated two major ways

²⁶ The development of a masculine *ǰa:* as opposed to PWS **ǰi:* in Dadanitic can be explained by four-part analogy to the inherited PS feminine *ǰa:t*, where the final *-t* is understood as the feminine marker: BNT ‘girl’: BN ‘boy’ :: *ǰa:t* : X = *ǰa:*.

Table 5.6: Potential sites for analogous extension of gender and number marking, based on forms common to most modern Arabic dialects (W. Fischer and Jastrow, 1980, pp. 61-3, 79-80). Apparent stems are shown to the left of the vertical lines, apparent formatives to the right.

2P pronouns		3P pronouns		2P PST verbs		2P PRES verbs	
M.SG		(a)		u(wwa)		a	Ø
F.SG	ʔint	i	h	i(yya)	katabt	i	i(:n)
M.PL		u(m)		um		u(m)	u(:n)
F.PL		in/an		in		un/in/an	i(n)

of marking gender. One approach was to innovate new vowels to mark gender, based on paradigms that were widespread in Arabic and where similar vowels mark similar qualities of masculine, feminine or plural. Table 5.6 shows how speakers could have interpreted certain pronominal and verbal endings as operating as morphemes or formatives which they could have used to build the new, gender marked demonstratives masculine $\delta a:$ and feminine $\delta i:$, what I will call VOWEL-ALTERNATING DIALECTS.²⁷

An originally genderless demonstrative $*\delta a:$ could also have been extended by creating a form $ta:$, by analogy with the various feminine formatives marker with t what I will call CONSONANT-ALTERNATING dialects. All of the dialects in the data examined here which have consonant-alternating demonstratives have the same vowel in both the masculine and feminine forms, that is to say there are no demonstratives forms which are both consonant and vowel alternating. If these dialects had inherited forms that were gender differentiated with m./f. $\delta i:/\delta a:$ we

²⁷ Huehnergard (2006, fn. 61) notes that for the det.-rel. pronouns, a similar leveling process appears to have begun in Ugaritic, OSA and Ethiopic, where the masculine singular took over the function of the feminine singular and plural det.-rel. pronouns.

would expect to have find dialects of a type $\delta i:/ta:$ which do not occur in the data. Indeed, such dialects would not have a particularly strong motivation to innovate $ta:$ if they already had gender marked singular demonstratives.

We do not even find $\delta a:/ti:$ dialects which we might expect from contact between the vowel-alternating dialects and the consonant-alternating dialects which are in close proximity to one another as seen in Figure 5.1.²⁸ Were this a leveling process, where the vowel was leveled since there was no ‘need’ to doubly mark the gender distinction, we would still expect to find more remnants of the early stage with unbalanced vowel marking.²⁹

What we do find is dialects of the type $\delta i:/ti:$, but this seems to be a later process of word final vowel raising, especially given intermediate dialects such as Behnstedt and Woidich’s (1985) Yemen point #37 where the near singular demonstratives are $\delta e:/te:$ but the distal forms are $\delta a:k/ta:k$. In fact, all of the consonant-alternating dialects in Yemen have a low vowel in the distal form, where the vowel is not word final, regardless of the vowel quality in the proximal forms (see especially Behnstedt and Woidich, 1985, maps 51-54). Moreover, there is a clear, smooth development from low vowels, to mid-vowels (as in Yemen #37) to high vowels, all suggesting a purely phonological raising process as opposed to e.g. the leveling of a high versus a low vowel.

²⁸ There may be an exception to this in the southwest Saudi Arabian city of Abha, though the data here is not very clear, since Theodore Procházka Jr. (1988) reports a form for the masculine distal demonstrative of $\delta a:k$ for that city, but $ta:k$, $ti:k$, $tiya:k$ as the feminine distal marker, though it is not clear that these come from the same speakers or even the same dialect per se.

²⁹ Al-Jallad (2012, pp. 316-319) derives the demonstrative $ta:$ from a reanalysis of the interrogative he reconstructs as $*matay$ ‘when’ as $ma-tay$ on the basis of $ma:-\delta a:$, but he is still left with the problem of how this replaced the feminine demonstrative. Moreover, the vowel is clearly originally low, and not a diphthong as discussed in this section. Finally, the attested modern consonant-alternating dialects do not generally have a reflex of $*matay$ for their temporal interrogative, and show quite a bit of diversity (Behnstedt, 1985, map 66), though I acknowledge that

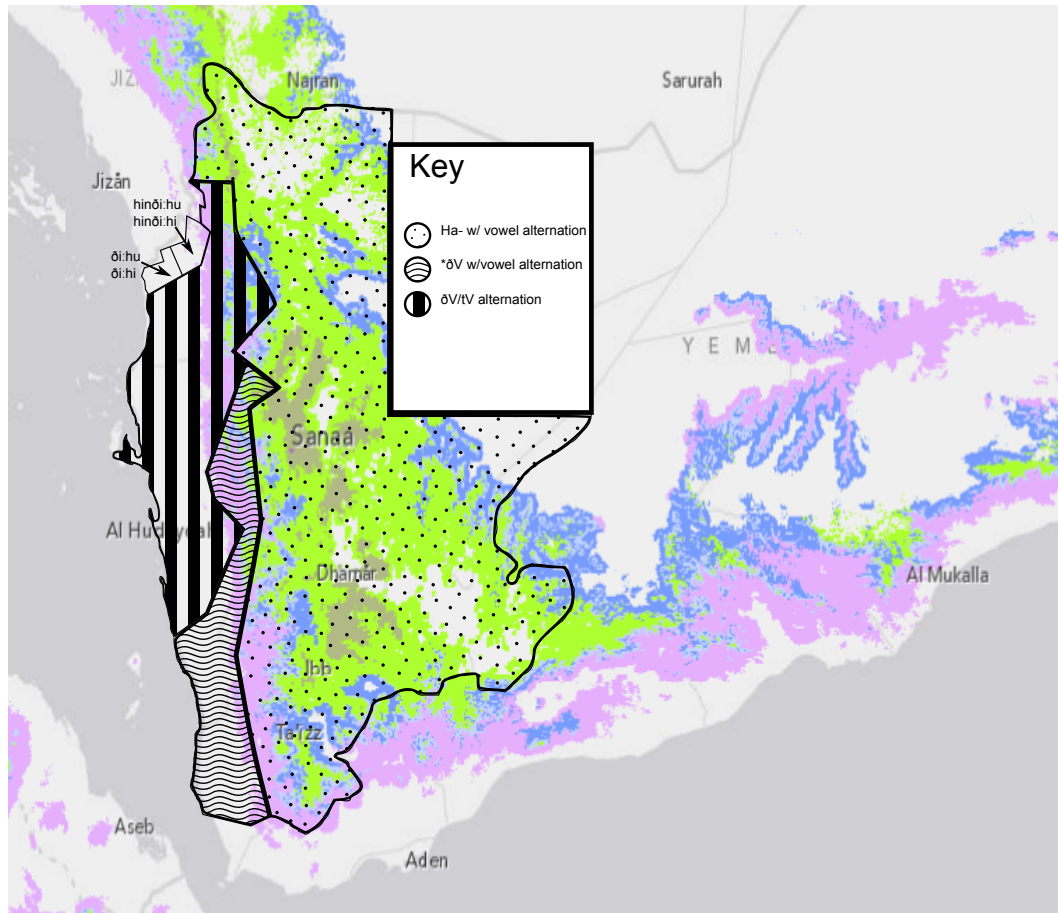


Figure 5.1: Yemeni proximal singular demonstratives, divided according to whether they are of the type haðV or ðV (i.e. with no ha- prefix) which use vowel alternations to mark gender, or have an alternation between ð/t to mark m/f gender respectively.

There are also a number of dialects which show gender-neutral use of **ḏa*;³⁰ that is they retain in some cases and phrases the archaic genderless use of that demonstrative. W. Fischer (1959, p. 57) gives examples of the following phrases in the dialect of Tetouan: *da-r-raʒəl* (ḏa-def-man) ‘that man’, *da-l-mra* (ḏa-DEF-woman) ‘that woman’, and even *da-rʿ-rʿʒəl* (ḏa-DEF-men) ‘those men’. He gives similar examples from other dialects particularly those in Yemen and Morocco (see also *ibid.*, pp. 58-9), with some from the dialect of Dofār in Yemen using a higher vowel, though again this is probably due to word-final raising. In all of these cases, the vowel seems to be an original *-a:.

There may also be some evidence for the *ḏa*: only stage in Quranic (and Classical) Arabic, where the interrogatives *ma*: ‘what’ and *man* ‘who’ are followed by an element *ḏa*:, a structure which according to Huehnergard and Pat-El (2007, p. 332) represents “frozen vestiges” of an old clefting pattern, with *ḏa*: acting as a relative pronoun.³⁰ In Classical Arabic, *ḏa*: is normally treated as the accusative form of **ḏa*:-type demonstratives, especially those operating as det.-rel. pronouns. However, in these cleft constructions, the fronted element is not necessarily extracted from an accusative position (*ibid.*, p. 328), and instead the vowel seems to be a frozen form. Given the relative archaism of the construction, and the fact that the vowel is /a/, this suggests that the construction may have been formed with the only singular demonstrative form available at the time, *ḏa*:.³¹

this could be due to later changes.

³⁰ They suggest also an identity between this form and the *ḏa*: in *ha:ḏa*-type forms (p. 333), a claim we will discuss below.

³¹ An alternative possibility exists, that this is simply a masculine form of the demonstrative, which could indicate the phase of Classical Arabic that gave rise to the pseudo-cleft *ḏa*: could have already differentiated to become either a consonant or vowel alternating dialect, since we don’t see a comparable feminine form. There are some traces, however, in the Baskinta Lebanese dialect where we have interrogatives such as *laf-ta* (why-ta) or *ki:f-ta* (w)ith the feminine, consonant-alternating form, a possibility that could have existed in earlier varieties of Arabic that were

The later form which largely replaced this relative particle in Classical Arabic is (m., f., c. pl.) *ʔalla-ḏi*, *ʔalla-ti*, *ʔalla-ḏi:na*³² Here the paradigm of the attached **ḏa*-type demonstratives very closely resembles that of consonant-alternating dialects, and the plural demonstrative is characteristic of many Yemeni dialects. Indeed, though no modern dialect in our sample shows all of these forms in a single paradigm, it would not be surprising at all to find an ancient dialect of the type where the (proximal) demonstratives were (m., f., pl.) *ḏi*, *ti*, *ḏi:n(a)*, with the singular demonstratives found in dialect such as Yemen #37 discussed above, and the plural form widespread somewhat further south, e.g. in Yemen #157 where the proximal demonstratives are *ʔaḏa*, *ʔaḏi*, *ḏi:ne*.³³

The vowels in these relatives are often considered to be a reflection of earlier case marking, but their resemblance to the demonstrative forms common in consonant-alternating dialects are striking, and the evidence above suggest that the vowel in these forms are the result of a phonological process and not a leveling of a certain case-form. This would suggest that there was no case-marking function per se for the endings of the relative marker in Classical Arabic, but instead that these endings were simply derived from a demonstrative paradigm that became suffixed to the form *ʔalla*-.³⁴

included in Classical Arabic.

³² Often presented as the definite article *ʔal*- followed by a somewhat mysterious *la* (see for example J. Barth, 1913, §44c), often treated as an ‘asseverative particle’ per Huehnergard (1983), though there is rarely a principled syntactic explanation given for how these forms were combined. Moreover, it is not clear whether these elements would have been analyzable as separable morphemes by the stage of Classical Arabic by speakers.

³³ The proximal forms aren’t certain, as the map has a typographical error, but it seems like this is likely the correct forms, or at least forms which are clearly derived from **ha:ḏa/i*.

³⁴ Al-Jallad (2012, p. 89) repeats Rabin’s (1951) (p. 89–90) assertion that some variants (the name given by Rabin is that of the tribe Huḏayl) of Classical Arabic had case marking in the plural, with *ʔallaḏu:na* and *ʔallaḏi:na* for the nominative and the oblique respectively. However, this claim seems to be based on a misinterpretation (or perhaps elaboration) of a statement by the grammarian al-Farrā’ ([d. 822] 1983, v. 2 p. 184) regarding verse 20:63 :

The split into vowel-alternating and consonant-alternating dialects appear to have been a very old split, and a very abrupt one — intermediate forms, as discussed above, are almost non-existent. It is primarily the dialects with vowel-alternating forms participated in the later changes which resulted in the fusion of the **ha:* presentative marker and the **ǧa:* demonstratives (discussed below in Section 5.4.2.1).

There is a small number of dialects which do have *ha:tak*,³⁵ but these are mostly Yemeni dialects on the very southern border of Northern Yemen, and hence on the border of the maps in Behnstedt (1985), so it is unclear what the form of demonstratives is in the dialects to their immediate south. However, it should be noted that these dialects have vowel-alternating demonstratives for their proximal

كما قالت العرب (الذي) ثم زادوا نوناً تدل على الجماع، فقالوا (الذين) في
رفعهم ونصبهم وخفضهم كما تركوا (هذان) في رفعه ونصبه وخفضه. وكنانة
يقولون (اللذون)

As the Arabs say, *ʔLDY* and then added an *-n* to mark the plural, so they
use *ʔLDYN* in the nominative, accusative and genitive, just like they use *HDʔN*
[this.DUAL.(NOM)] in the nominative, accusative and genitive. And [the tribe of]
Kināna says *ʔLLDWN*.

Here al-Farra is clearly saying that this is invariant form, i.e. that it does not decline for case. Claims that this form actually did inflect for case probably came much later, reflected in Rabin's usual use of very late sources for this data — almost all of his sources for this are from the thirteenth century CE. Interestingly, the Kināna tribe appear to have dwelt near Hudayl in pre-Islamic times, so the confusion is not surprising given the centuries that passed between al-Farra's account [d. 822] and these later sources. It is also one reason that I am reluctant to rely too heavily on Rabin (1951).

³⁵ Yemen points #104,119,130,140, 23.

series, and consonant-alternating demonstratives in their far series. The far series demonstratives in Arabic tend to preserve older demonstrative forms. A nearby dialect, in Yafi^c (200 km NNE of Aden) is said to be losing its original consonant-alternating demonstratives to vowel-alternating demonstratives (Vanhove, 2004), probably as part of the larger process of areal dialect leveling that is affecting other parts of the grammar (Vanhove, 1995, p. 151).

This evidence suggests a direction of change where originally consonant-alternating dialects are now borrowing the vowel-alternating demonstratives, which enter first into the proximal demonstrative series, probably due to being more frequent, and hence more salient in the contacts between speakers of the dialects involved. This might explain forms such as *ha:ta:k* where the **ha:* form was borrowed along with the near demonstrative series, and then transferred to the far series (which could have consonant-alternating forms) by analogy. An example paradigm that shows the general lack of balance in the proximal and distal sets in these dialects can be seen in the following table:

Table 5.7: Demonstrative forms in Yemen #130.

	m. sg.	f. sg.	m. pl
Proximal	ha:ða	ha:ðih#	ha:ðawla
Distal	ha:ða:k	ha:ta:k	hawla:k

5.3.1.2 Diminutives

Originally, some dialects appeared to have had both a normal series of demonstratives as well as a diminutive series. The function of the diminutive series is not exactly clear, but Diem (1973) refers to these forms as “verkleinernd”

i.e. diminutive or “verächtlichend”, i.e. derogatory. Similarly, in the treatment in the earliest Arabic grammatical work, Sībawayhi (1999, v. 3, pp. 540-543), these forms are referred to as *muḥaqqar* which, while clearly used as a technical term, might be translated as ‘made contemptible’, though elsewhere he does refer to related forms as *muṣaḡḡar* ‘diminutive’. Brustad (2008) gives a thorough overview of some of the functions of diminutives across languages, and in Arabic, which includes the use of diminutives to express contempt.

The only use of these forms I could find in Diem’s (1973) texts was at the culmination of a story where a man has been forbidden from eating his own lambs, but ends up eating with someone who serves him the forbidden meat. Right as he begins to eat, the storyteller says:

(7) (Yarīm dialect, Yemen Diem, 1973, 134, ln. 16)

sa:r yi-tyadda ſind ḏayya:k ḏi ftara al-ʔatʔla
go.PST.3S 3S-eat.lunch at that.M.S.DIM REL bought.3S DEF-lamb

He went and ate with that [(cursed) fellow] who had bought the lamb.³⁶

In the context of this story, the diminutive seems to emphasize the almost evil nature of the person who is giving him the forbidden lambs, and would indeed seem to fall under the rubric of making him contemptuous.³⁷

Morphologically, these forms are formed by adding a suffix *-yya* to primarily the singular demonstratives, producing forms such as masculine *ḏayya*, feminine *tayya*, and apparently in the Iryān dialect in southern Yemen, plural *ḏaulayya*

³⁶ Original: “Er ass also bei dem, der die Lämmer gekauft hatte.”

³⁷ This is not simply the normal demonstrative form — cf. line 10, *ha:ḏa θ-θawr* (this DEF-bull) ‘that bull’ or line 11 *haḏa:k al-wulaid* (that DEF-boy.DIM) ‘that (small) boy’.

(Diem, 1973, p. 48).³⁸ This suffix is probably related to the *ay* formative that is part of the widespread Semitic *qutayl* diminutive pattern, and which is also used word-finally in Ethiopic languages, though based on my information, only Arabic has these forms in the demonstratives (Brockelmann, 1928; Zewi, *EALL* : “*Diminutives*”).

Diminutive demonstratives are discussed in Sībawayhi (1999, *bāb taḥqīr al-’asmā’ al-mubhama*, vol. 2, 540-543), who states that diminutives can be formed from all items derived from **ḏa*:-type demonstratives, including the Classical Arabic relative pronouns. However, he claims that “they[Arabic speakers] avoid making a diminutive of the feminine form *ha:ḏihi*, which would be confusing”³⁹, presumably since it would result in *haḏiyya*, though there is no reason to believe that such a form would truly result in a homonymic clash. Instead, this may simply be the result of the fact that these diminutive forms tend to be more prevalent among consonant-alternating dialects, where the feminine forms are *tayya*.⁴⁰

Distal forms are derived from the proximal forms, so we find *ḏayyak* type forms (also reported in *ibid.*). Some dialects (e.g. in NW Yemen) show *ḏiyya*, *tiyya* with an /i/ in the first syllable, but this is probably due to the influence following glide raising the vowel.

In a number of dialects, through a process of markedness reversal, these diminutive demonstrative forms have become the primary demonstratives. Other dialects may have both normal demonstratives as well as diminutive demonstra-

³⁸ See also Yemen #23, which has plural *ḏa:liyah*.

³⁹ > > وكرهوا أن يحقروا المؤنث على هذه فيلتبس الأمر < <

⁴⁰ It is interesting to note that the poetic examples in my edition of Sībawayhi’s chapter on diminutive demonstratives have been ‘cleaned’ so that they do not actually show the diminutive forms, though I believe I did see the same verses cited in some other grammar with the diminutive forms intact.

tives, which could have fallen out of use. It would seem based on the text from the Yarīm dialect that the diminutive forms may be a type of relic in that dialect, since the form has no **ha:-* prefix while the primary demonstrative series in that dialect do, suggesting that the diminutive is a frozen older form, not a form that is synchronically derived from the modern demonstrative series.

There are dialects in North Africa which show forms such as *ha:ða:ka:ya* (Tripoli, W. Fischer, 1959, p. 99), which are said to operate as emphatic forms, but the position of the *-a:ya* suffix suggests a much later process of deriving these forms, probably based on the emphatic pronouns in these dialects which also show an *-ya* suffix.

5.3.2 Development of plurals

5.3.2.1 Plural forms with /l/

The plural demonstratives (derived from either **ʔl* or from **ʔlV*) were also originally not differentiated for gender, though the inherited vowels in these forms are far from clear. As Hasselbach (2007a, p. 15) notes, it is not clear whether the proto-Semitic vowels in these demonstrative forms were originally /u/ or /i/ vowels, and we find both in Arabic dialects, in addition to forms which may have originally had an /a/. Hasselbach (*ibid.*, p. 23) prefer a unitary explanation with /u/ since /i/ developments could be seen as being specific to Hebrew and Aramaic: the second syllable of their demonstratives is *-li*, and to avoid the impossible nominal form *quttiḷ*, the initial syllable was changed by vowel harmony. Based on the distance between Arabic and Akkadian, which she both classifies as having /u/, she suggests an original /u/. However, Arabic (as she herself notes in fn. 49) does have forms with /i/. Moreover, the final vowel in many Arabic dialects is /a/, not /i/ so a vowel-harmony type explanation is also not necessarily productive. The

diversity in both Arabic and Semitic argues instead for an assumption of earlier diversity, with both /u/ and /i/ forms co-existing in Semitic.⁴¹

Forms which clearly derived from both an original *ʔul(l)a* and an original *ʔil(l)a* are widespread.⁴² Examples are shown in Table 5.8 of Yemeni dialects which show representative examples of these forms.

Yemen #109 *hilla* strongly suggests derivation from **ha-ʔilla*, whereas Yemen #93 f.pl. *halla* is ambiguous — it can conceivably be from the same source, with the /a/ vowel of the prefix taking precedence over the /i/ vowel (or over diphthongization) possibly due to paradigmatic pressures, or from an original **ha-ʔalla*. Similarly Yemen #37 *ḍawli* strongly suggests a derivation from *ḍa-ʔuwla*⁴³, suggesting the existence of an original *ʔul(l)a*.

From these and other forms, we cannot say whether there was some original *ʔal(l)a* form, but given the Semitic evidence, and the lack of clearer attestations we will have to assume that such a form did not exist, but is the result of replacement by the preceding /a/ of the prefixed **ha:-* and **ḍa:-*.⁴⁴

⁴¹ It is difficult to ascertain whether there was any meaningful variation between these forms, especially whether they were used to distinguish masculine and feminine plural. The Semitic evidence simply is not very good on this point, especially given the relative infrequency of marking a gender difference in plural demonstratives. Most Semitic languages which distinguish gender in plural demonstratives actually do so using the forms which are identical to the pronouns, so this is of little help.

⁴² It is not clear whether or not the initial glottal stop was intact as a phoneme in any of the dialects discussed here at the time the demonstratives developed, but I will continue to use the notation with the glottal stop in a nod to the Semitic reconstruction. Glottal stops word initially preceding a vowel have an ambiguous status even in Classical Arabic (e.g. the difference between *hamzat al-qatʿ* and *hamzat al-waṣl*).

⁴³ The final /i/ would likely be a result of the word-final raising discussed above for other demonstratives, as seen by its absence in a CVC context. The lack of raising in the final vowel in feminine plural proximal *ḍayla* might be a result of that demonstrative being added to the paradigm after this phonological rule was active, since . A (relatively) recent borrowing would also make sense given the very different form of the distal counterpart, *ḍallak*.

⁴⁴ Such a form would, however, be of interest in that it would be quite similar to the initial element

Table 5.8: Plural demonstrative forms derived from **ʔVl(a)*

	m. pl. prox	f. pl. prox	m. pl. dist	f. pl. dist
Yemen #109	hilla	—	hillak	—
Yemen #93	hawla	halla	hawlakkah	hallakah
Yemen #37	ǧawli	ǧilla	ǧ ^s awlak	ǧallak
Yemen #6	ʔo:li	ʔe:li	ʔolak	ʔe:lak

The development of the diphthongized forms is also somewhat complex, and it is important to understand since these forms appear to be rather widespread. For example, Nile Valley dialects almost all appear to be derived from dialects with masculine and feminine plurals of the type *do:l/de:l* (< **ǧa:-ʔula/ʔila*).

The demonstrative forms in Yemen #6 would seem to suggest that diphthongization in these forms was original, followed by attachment to either **ha:* or **ǧa:* demonstratives. However, there is no clear motivation to create diphthongs from /u/ or /i/ in these dialects. It seems likelier that the diphthongization is a result of the suffixing of the plural demonstratives to forms with final /a(:)/ sounds, which then produce the diphthongs /aw/ and /ay/ (which in Arabic dialects often monophthongize to /o:/ and /e:/). Yemen #6 (in the extreme NW of Yemen) has the only demonstrative form in Northern Yemen with no initial **ha:-* or **ǧa:-* affixed to the demonstrative, and is surrounded to the south by *ǧayla*-based forms (for common gender) and to the east by *hawla*- and *ǧawla*-forms. Assuming the dialect originally did not have these prefixes, its contact with these dialects may well have led them to change original masculine/feminine plurals *ʔula/ʔila* to *ʔo:li/ʔe:li*

in the relative markers found in Classical Arabic and many Yemeni dialects of the type *ʔalla-ǧi*, and Rubin (2005) argues that the definite article in Arabic is actually derived from a plural demonstrative.

whether through misinterpreting the morpheme boundary as not *ha-wla* or *ḏa-yla* but as *h-awla/ḏ-ayla* and adjusting the demonstratives accordingly.

A more mysterious development is the association of forms with initial /aw/ (always a diphthong or derived from a diphthong) with masculine plurals, and /ay/ or /a/ with feminine plurals (though probably both are derived from the *?ila* demonstrative). In many of the Yemeni dialects either one form or the other is used for the common plural. This could just suggest that initial dialect variation was inherited and that dialects which have the contrast have simply remapped the dialect variation into gender variation.

Alternatively, however, there could have been two phases: First, variation in early dialects between *?ula* or *?ila* would have been reinterpreted as gender variation, perhaps in analogy with the singular third person pronouns *hu(wwa)* ‘he’ *hi(yya)* ‘she’ respectively. Second, as often happens in modern dialects, the descendants of these dialects generalized either of the forms as their common plural form. In that case, dialects which still make this contrast would be seen as relics, whereas in the first hypothesis those dialects would be later innovations

The two-phase is more complex, but there may be evidence that previously there was greater diversity in the demonstratives, which has subsequently been leveled. For example, though generally dialects with demonstratives with original **?ula* suffixes occupy the Yemeni highlands, while demonstratives with original **?ila* are found in the central Tihama littoral, we find at least one dialect (Yemen #84) on the coast which has *ḏawli/ḏayla* for the masculine and feminine plurals respectively. All of the neighboring dialects have *(ha)ḏayla*. Moreover, we have strong suggestions historically that the Egyptian dialects derive from Tihama dialects, which would be an amazing coincidence if the distribution of dialects which marked plural gender with *ḏawli/ḏayla* was as restricted as it is today.

There is, however, the fact that most of the dialects which mark gender by contrasting **ʔula* and **ʔila* in Yemen are sandwiched between those two major dialect areas, so it is a distinct possibility that this contrast is a relatively more recent development that post-dates any generalization of these forms. Given the evidence available to us today, it is hard to say anything determinate here, though it does seem likely that the gender contrasting dialects were more widespread on the Tihama littoral than they are today.

Many of these plurals become suffixed to either **ha:* or **ǵa:*, producing forms such as Yemen #93 *hawla* or #37 *ǵawli*. The **ha:* forms are clearly the product of the same process that produced **ha:-*ǵa:-*type demonstratives, treated below in Section 5.4.2.1.

The plural forms affixed to the **ǵa:-*stems, on the other hand, cannot be explained through the same process of juxtaposition followed by affixation, since we have few records of any kind of structure in which a singular demonstrative is immediately followed by a plural demonstrative. The most likely segmentation of the singular demonstratives in a dialect with *ǵa:*, *ǵi:*, *ʔila* would have been to as *ǵ-a:*, *ǵ-i:*, so we actually don't expect a levelling-type approach to produce *ǵa+ʔila* > *ǵayla*. This strongly suggests that there would have been contact with **ha:-*prefixing dialects, where the *ǵa:* could be seen as equivalent to *ha:* and thence extended into the plural paradigm, producing the diphthong in *ǵayla*.

A contact explanation would also explain the divergent vowels in these *ǵ-*initial forms, which vary between *ǵayla* and *ǵa:la*. If they misinterpreted forms such as *hawla* as *haw-la* and replacing the apparent **ha:-*part with *ǵa:* they would have produce *ǵa:la*, a form which could also actually be derived from *ǵ-awla*. If they interpreted it as *ha-wla*, they would produce the similarly attested *ǵayla*, or *ǵawla*

5.3.2.2 Other Plural Forms

There are a number of other plural forms whose origins are much more obscure, with some exemplary dialects and their forms listed below:

Table 5.9: Proximal demonstratives with final /n/-sounds in plural

	m. s. prox	f. s. prox	m. pl. prox	f. pl. prox
NADJ	ha:da	ha:di	ha:du	—
Yemen #158	ʔaða	ʔaði	ʔaðum	ʔaðin
LvS281	ha:ða	ha:ði	haðo:l	haðanna
NABEN	ʔaðʕa	haði	haʕðo:l	haðe:n
Yemen #145	ha:ða	ha:ði	ha:ðum	ha:ðe:n
Yemen #156	ðe:	ði:	ðinne:n	

The simplest forms to explain are those represented by NADJ, where *ðu is a clear example of the same kind of analogy that created the /a/ (masculine) vs. /i/ (feminine) distinction, probably derived from the pronouns (*ʔanta/ʔanti/ʔantu(m) ‘You (m.)/You (f.)/You (pl.)’), but with similar contrasts throughout the grammar, such as the past-tense verbs *katab(a)* ‘He wrote’ vs. *katabu:* ‘They wrote’. This analogous form would have supplanted the original ʔVla demonstratives, so that dialects which have this base do not generally have bases of the type -ðawl in the plural.

A similar process is responsible for the development of the forms represented by Yemen #158, where a masculine form ending in -um is contrasted to a feminine form of the type -in. In the Arabic pronoun system, both /u/ and /m/ are taken to represent masculine plurals, while /i/ and /n/ can be taken to represent feminine plurals, though in many dialects that make these contrasts, either the vowel or the nasal changes while the other does not, so in SHUK the 3rd plural pro-

nouns are *hun/hin* for m/f respectively, and in ARQAH they are *hum/hunna*. The 3rd person plural pronouns in Yemen #158 are, of course *hum/hin*. This also may explain the first syllable in Yemen #156 *ḏinne:n*, which would have been extended based on an analogy along the lines of *inti:ḏi::intin:X=ḏin*, and then doubly marked later on, much like Yemen #150 which shows m., f. plural forms *ha:ḏin*, *haḏinnin*.

The other forms are much more difficult to account for. There is a superficial resemblance between LvS281's far demonstrative (derived from the near shown here) *haḏannitf* (with /tʃ/ < /ik/) and Classical Arabic nominative dual *ḏanika*.⁴⁵ Hasselbach (2007a, fn. 43) suggests that the Classical Arabic form is derived from **ḏa:nlika* with nC > nn assimilation, but there is no indication from whence this /l/, especially given the complete lack of /l/'s except from etymological /*ʔVl/ in dialectal Arabic.⁴⁶ As discussed previously, there are almost no examples of surviving -l- in modern dialects equivalent to that in Classical Arabic *ḏa:lika*, so this is also an unlikely explanation for the modern dialect form.

This form could be the result of contact. Syriac, an Aramaic dialect, had an alternation between a common proximal plural *ha:le:n* and a masculine distal plural *ha:nnon* and feminine plural *hanne:n* (see *ibid.*, p. 14). In this scenario, the Arabic distal feminine plural demonstrative would have been something like **haḏallak*, with masculine plural **ha:ḏo:la:k*. Speakers might have understood the form *haḏallak* as comparable to with *hanne:n* in some way, and the /ll/ would have become /nn/, producing *haḏannak*, or even potentially, with a raising of the final vowel on the model of *hanne:n*, which would explain the disjunct between the vowel of LVS281's distal feminine plural *haḏannitf* versus the proximal *haḏanna*.⁴⁷

⁴⁵ Jarbou, 2012, 111, calls this -nn- form a Hijazi demonstrative, versus "Tamimi" *ḏa:nika*.

⁴⁶ The irregularity of nC > nn assimilation, which Hasselbach admits, is particularly confusing given Classical Arabic's preference for nC > CC assimilation, widespread in the Quran for example.

⁴⁷ This is ultimately not a very satisfying explanation. I am open to other possibilities.

This form also could simply be the result of a fluidity which exists in Arabic dialects between /l/ and /n/, seen in a number of words: In the Levantine we hear both *mli:h~mli:h* ‘good’ and *bortu?a:l~bortu?a:n* ‘orange’ (Behnstedt, 1997, Map 352, 450), while in Egypt we have *fiŋga:n~fiŋga:l* ‘coffee cup’ (Hinds and Badawi, 1986, p. 672). Similarly, /l/ and /n/ are fluid throughout Semitic — Moscati et al. (1964, p. 32) give numerous examples of similar couplets, such as Akkadian *lamṣatu~namṣatu* ‘fly’ and Phoenician BL ‘son’ for BN.⁴⁸ That there was a shift to /n/ from /l/ and that it was normalized is of course very diagnostic for these dialects, because it is such an unusual shift, even if its provenience is unclear. Without clearer evidence of contact, I will assume that this form is just the result of an ad-hoc sound change.⁴⁹

The forms with /-e:n/ appear at first to be related to Classical Arabic oblique dual forms with /ayn/ (which often > /e:n/). However, the phonology of the dialects with this /-e:n/ do not suggest that this is the case. Many of the dialects which show /-e:n/ still have the diphthong reflex of /ay/, including Yemen #156 (Behnstedt, 1985, compare map 12 with map 52). Instead, a different phonological process seems to be happening here: in most of these dialects, /in/ and /un/ are lowered and lengthened word finally before a pause, producing /e:n/ and /o:n/ respectively, that is before a coronal nasal /n/ (ibid., map 16).⁵⁰ This would explain Yemen #145’s forms as being actually derived from the same forms as nearby #165’s *?aḏum/?aḏin*. Moreover, a note on map 52 (ibid., p. 20) notes that *ḏe:ne* is actually a

⁴⁸ There is also the well known Aramaic shibboleth of BR for ‘son’ which is quite similar to the *ḏanna* demonstratives here in that it is an ad-hoc change, but that ad-hoc nature makes it very diagnostic of a particular linguistic group.

⁴⁹ An intriguing parallel to this is the difference between the word for ‘statue, idol’ in ANA and Arabic: In ANA, the word is *šlm* while in Arabic it is *sʿanam*. See further Scagliarini (2007).

⁵⁰ Apparently /a/ also gets lengthened in -aC# pausal contexts, but apparently with no change to the vowel quality. Behnstedt (ibid., p. 15) with his usual concision: “Parallel zu *i,u* kann auch *a* in -aK# gedehnt werden.”

pausal form, appearing as *ði:ne* in context, again suggesting phonological origins for these apparently unusual forms with /e:/. These pausal examples with /i:/ > /e:/ before a nasal (but not word finally) also suggest that the phonology of these dialects is more complex than it would appear simply from map 16.

5.4 The evolution of **ha:*

Almost all Arabic dialects, including Classical and Quranic Arabic, have an element **ha(:)* which takes on various functions, and indeed related elements are widespread throughout Semitic. As argued above in Section 5.2.2, the element **ha:* shares some developments, but not all with other Semitic languages, and in this section we will largely deal with the specific case of **ha:* in Arabic.

The goal of this section is to understand how the many, related functions of **ha:* arose. In some dialects of Arabic, such as Moroccan Arabic, **ha:* is a free, general presentative particle, as in (11). In most dialects it appears to either be affixed to the demonstrative as in Classical Arabic *ha:ða* ‘this’. In many dialects, especially in the Levant, **ha(:)* has gained gender and often number inflection, creating a new series of adnominal demonstratives, though they can also be a pronominal as in LvSou *ha:* ‘this (m.sg.)’.

5.4.1 Original function of **ha:*

In this section, I will argue that the basic meaning of **ha:* was as a general, attention gathering particle, which developed into both a referent presentative and a sentential presentative. This process happens in many different Semitic languages suggesting that this **ha:* arose quite early in Semitic, but many of the later developments of **ha:* into other functions tend to be language specific.

The status of **ha:* as some sort of attention gathering particle is not controversial. The early grammarian al-Mubarrad (*al-Muqtaḍab*, p. 275) [d. 898], describes it as a *ḥarf at-tanbīh* ‘particle for calling attention’, and W. Fischer (1959) calls it a “sentence-introducing particle” (Satzinleitungspartikel), whose function he describes:

Als demonstrative Interjektion bedeutet *hā* zunächst etwa “siehe (dahin, da, dort),” es soll den Angesprochenen auf die folgenden Aussage aufmerksam machen oder deren Wichtigkeit hervorheben.

As a demonstrative interjection, *ha:* expresses primarily something like, “look (here, there, over there)”, it may call attention to the following statement or to emphasize its importance. (pp. 157-158)

Fischer’s analysis, however, implies an endophoric function, with **ha:* marking the relative importance of sentences or statements. However, the original function of **ha:* is extra-textual, i.e. exophoric, that is to say, it simply calls for attention from the hearer, which will of course only precede sentences whose content is important. To extend the signpost analogy of Bühler (1934) from Section 5.1, **ha:* does not even have an arrow pointing to what attention is being drawn toward — instead, it is equivalent to a sign labeled “Attention!”, with the onus on the hearer to figure out what to pay attention to. This explains usages such as the cry of Jewish Fessi vendors *ha: ssmaʃ* (ha: listen.IMP.SG) ‘Hey! Listen!’, where there is no textual function, but only the attention grabbing function of the cry. The source of the voice provides the necessary clue about where attention should be directed.

It is necessary to make an original appeal to this exophoric function in order to explain the fact that **ha:* is the source of vocatives in some varieties of

Arabic, and in Ancient North Arabian (ANA), the most closely related Semitic language to Arabic.⁵¹ In Thamudic B, a variety of ANA attested throughout the Arabian peninsula, *h* is attested as a vocative particle, as well as in a demonstrative use (Macdonald, 2008a). In SHUK, *ha:* is a vocative form still, as in *ha: walad-i* (ha: son-mine) ‘Hey, my son!’ (Reichmuth, 1983, p. 195).⁵² Similarly, in many North African dialects, the vocative *ʔa* appears to be derived from **ha:*, related to the same ad-hoc sound change that appears to have created the Egyptian presentative *ʔa-* (W. Fischer, 1959, pp. 166-167).⁵³

5.4.2 **ha:* as a presentative

The leap from a general, attention gathering particle to a presentative form is not surprising, since speakers will tend to call for attention specifically before indicating something new or important. One such function is an exophoric pointing function, as in (8), where the sentence would be meaningless without some sort of gesture. The particle **ha:* is also frequently used to show that something is available, close to the subfunction of presentatives as offeratives, as in (9).

(8) (Kuwaiti Arabic: from Brustad, 2000, p. 124)

rəħt ʕind l-galʕa:f sawwa:-li kərsi ha tʕu:l-ah
 went.I to DEF-shipbuilder made.he-for.me chair ha: length-its

I went to the ship-builder and he made me a chair *this* high (lit. this is its height)

⁵¹ I recognize that this is a very subtle distinction indeed, but one which I find helpful conceptually.

⁵² Original translation: “Ach, mein Junge!”

⁵³ Indeed, the change of /h/ > /ʔ/ is widespread in Arabic, with examples in the demonstratives as discussed here, as well as in the change from the Western Semitic verb form /haqtala/ to Arabic /ʔaqtala/ (i.e. Form IV).

- (9) (Mosul, Iraq W. Fischer, 1959, p. 163)

haʔ ʕafa:-ku
 ha dinner-your.PL

Here's your dinner

However, **ha:* can also be used to draw a listener's attention to a full sentence, as in (10). This use seems to be the basis of a less clearly exophoric-deitic meaning, and into a general function as a sentential presentative, as in (11). In that sentence, while it could be interpreted as presenting the mother-in-law — a general ambiguity that is the source of many of the later **ha:* developments — it is operating as Fischer claimed, to emphasize the import of the sentence generally.

- (10) (LvDAM, *ibid.*, p. 163)

ha: xud-i: hadi s-suwa:re
 ha: take-FEM.SG this.F DEF-bracelets

Take these bracelets

- (11) (Moroccan Arabic: Brustad, 2000, p. 122)

ha ʕgu:zt-i tnuww-at. ga:l-t ...
 ha: mother.in.law-my caught.on-FEM said.FEM

See, my mother-in-law caught on. She said...

It is unclear whether **ha:* began as a referential presentative, or as a sentential presentative. Though a sentential presentative might appear to be the more general use, the function of introducing new entities into the discourse is probably more relevant to everyday language use than gathering attention before starting a sentence. It is possible that **ha:* began as a referential presentative, as in (9) where

the phrase following **ha:* is incomplete and that phrase depends on **ha:* in some manner, triggering interpretations which could range from **ha:* as an offerative to a demonstrative adverb (e.g. “Here”). This is the very same tension as in English *Here is your dinner*, though *here* is clearly etymologically a demonstrative adverb, whereas **ha:* has a more nebulous role and it is only in this sentence that it could be seen as operating in this manner. In phrases like (12), the clause following **ha:* is complete, and could stand entirely by itself, with no need for the **ha:* that is found in the clause, so **ha:* could be reinterpreted as a sentential presentative. Or, alternatively, the two presentative functions — referential presentative and sentential presentative — are sufficiently similar that both would be a natural evolution of an attention gathering **ha:*.

The same lack of dependence on **ha:* is seen in (13), where the clause beginning with a pronoun and followed by a verbal predicate suggests a function of **ha:* as a sentential presentative, as opposed to the likely more original structure (14) where **ha:* presents an object, not a sentence.

- (12) (Casablanca, Morocco W. Fischer, 1959, p. 159)

ha: lba:bo:r za
 ha: steamboat come.PST.M

Look, the steamboat arrived or: Look, the steamboat, it's arrived⁵⁴

- (13) (Tangier, Morocco *ibid.*, p. 164)

ha-nti ka-t-fuf da:ba
 ha:-you IND-2-see now

⁵⁴ This could even be interpreted, as it probably was in dialects with a **ha:*-based adnominal demonstrative, as “This steamboat has arrived.” Note that (*ibid.*, p. 158) also gives the sentence *ha: ba:bo:r za:i* (ha: steamboat come.ACTPART.M) ‘Look, a steamboat has arrived or: Look, a steamboat, it's arrived’, where a demonstrative reading isn't as readily possible.

Now you'll see!

- (14) (Casablanca, Morocco W. Fischer, 1959, p. 164)

ha:-na:

ha:-1SG

Here I am [at your service].

5.4.2.1 Attachment to *ǝa:* forms

The function of **ha:* as a sentential presentative means that it could precede any type of sentence, functioning to emphasize the importance of that sentence. It could even precede clauses that were already presentative, and this is how **ha:* became attached to **ǝa:* in Arabic to form first a new sentential demonstrative, which then was reinterpreted as a demonstrative. In many dialects, the **ha:- *ǝa:* demonstratives finally replaced the original **ǝa:* demonstratives which they were originally formed from.

Like any pronominal demonstrative, **ǝa:* demonstratives could function as the subject of a sentence, which in many cases is an originally exophoric function, as in the following examples:

- (15) (EGCAI, Woidich, 2006b, p. 44)

da *kanz*
DEM.M.SG.NEAR treasure

This is a treasure

- (16) (EGCAI, *ibid.*, p. 44)

wi da *yi-sʃa* *fi ʔatʔi ʃe:f* *da*
and DEM.M.SG.NEAR 3.M-strives to cut bread DEM.M.SG.NEAR

And this [one] strives to cut off that one's livelihood⁵⁵

- (17) (EGCAI, Woidich, 2006b, p. 44)

do:l sitta:t il-ʕima:rʕa byithiblu ʕalafa:n-u
 DEM.PL.NEAR women DEF-building go.crazy over-him

The(se) women of this building go crazy over him⁵⁶

However, this essentially pronominal use of the demonstrative (but substituting for an exophoric rather than endophoric referent) is extremely similar to a presentational structure, and this indeed is one of the functions of **ǰa:* in dialects that have not fused **ha:* with **ǰa:*. In the dialect of Ūlād Brāhīm in Algeria, sentence initial use of **ǰa:* is only understood as a marked referent presentative function with affective meaning, as in (18). In (19), the purely presentational function of *da* is clear, especially from the lack of agreement of the demonstrative with the following noun.

- (18) Ūlād Brāhīm, Algeria (W. Fischer, 1959, p. 179)

ǰa: ʕæ:rəs
 DEM.m.s. horseman

Here's a good rider⁵⁷

- (19) EGCAI (Woidich, 2006b, p. 181)

?ahlan! da l-ʕisʕa:ba kulla-ha wisʕl-it
 hello! DEM.M.SG.NEAR DEF-gang.F all-3F.SG arrived-3F.SG

⁵⁵ Original: “und der bemüht sich, das Brot von dem wegzunehmen”

⁵⁶ Original: “die Frauen des Wohnblocks sind verrückt nach ihm”

⁵⁷ Original gloss: “en voilà un bon cavalier”

Hello! The whole gang is here!⁵⁸

The of the **ǰa:* demonstratives to introduce sentences would likely have existed very early in Arabic dialects, and so would have been inherited in probably all dialects. Similarly, since all of the ancestors of modern dialects allowed sentences without explicit copulas, the sentence introducing demonstrative would always have had the ambiguity to be interpreted as either a simple pronominal demonstrative in subject position, or as a sentential presentative or referent presentative. If this were the case, we would expect some examples where the **ǰa:* demonstrative series continued to grammaticalize accordingly. For example, Diessel (1997, 1999) argues that demonstrative identifiers, what I call referent presentatives, often grammaticalize into copulas, and we should expect to find such an example given the large number of Arabic dialects in which **ǰa:* had this dual function. Indeed, this process has occurred in the Yafi^c dialect of Yemen (some 200km NE of Aden), where the original feminine singular demonstrative has become a presentative, and then a copula as shown in example (20). Note that this is not the only function of *ta:* in this dialect, which retains its role as a demonstrative, though it is apparently being lost to a series of vowel-alternating demonstratives.

(20) Yafi^c Arabic, Yemen (Vanhove, 2004, p. 331)

u y-fqaff-ha: bi-r-ra:s u ra:s-ha: ta nos^f-e:n
and 3.M-hits-her in-DEF-head and head-her COP half-DUAL

He strikes her in the head, and her head is split in two⁵⁹

⁵⁸ Original, note the use of ‘ja’: “willkommen! Die ganze Bande ist ja angekommen.”

⁵⁹ Vanhove’s translation: “Et il l’a frappée sur la tête et sa tête s’est retrouvée en deux moitiés.”

The pragmatic and discourse importance of a presentative sentence would likely have attracted the use of **ha:* to the **ǵa:* initial sentences. Presentative sentences are focal points in the discourse, dictating what further speech will be about, and it is natural that an attention gathering, sentence modifying particle such as **ha:* would help to introduce such sentences. The particle **ha:* would clearly be a sentential presentative, while **ǵa:* was primarily a referent presentative, meaning that a juxtaposition of **ha:* and **ǵa:* would not be ambiguous.

We can therefore reconstruct the series of events which led the juxtaposition and fusion of **ha:* and **ǵa:* as follows (and as is shown in (21)): **ǵa:*-initial clauses and **ha:*-initial clauses might both be used to present objects in the world, but **ǵa:* was largely used as a referent presentative, while **ha:* was used largely as a sentential presentative (Stage 1). The two would be used together in presenting a sentence. The sentential presentative **ha:*, having no function in the clause, would have preceded **ǵa:* which acting as the referent presentative, operated as the subject of the clause (Stage 2). The frequent co-occurrence of the two particles would result in their fusion (Stage 3). This fused form would first operate as a sentence initial referent presentative. The ambiguity between an identificational demonstrative and an adnominal demonstrative would lead to a reanalysis of this as either a normal pronominal demonstrative, or even an adnominal demonstrative (Stage 4).

(21) Posited example

<i>Stage 1</i>	> <i>Stage 2</i>	> <i>Stage 3</i>
[ha:/ǵa:] [ar-rajul]	> [ha] [[ǵa:] ar-rajul]	> [haǵa] [ar-rajul]
PRESENT/DEM DEF-man	> PRESENT DEM DEF-man	> DEM DEF-man
> <i>Stage 4</i>		
> [haǵa ar-rajul]		
> DEM DEF-man		

Look, the man/This is the man > Look, this is the man > This (emphatic) is the man > This man...

One possible bridging context that would have allowed for the progression from Stage 3 to Stage 4 is shown in (22). Here speaker could either interpret *ha:* and *ði:k* as a single unit, i.e. ‘that is the money’ instead of the reading with them separate, ‘there is that [previously discussed] money’.

(22) (Ṣan‘ā (Sana’a), Yemen W. Fischer, 1959, p. 163)

ya:-ba: ha: ði:k al-filu:s bi-ða:t-ha:
oh-father ha: that.FEM DEM-money.F in-self-her

Oh Father, there is that very money.⁶⁰

The notion of emphasis in Stage 3 would have derived from the fact that *ha:ða:* and *ða:* would have both existed as referent demonstratives, but *ha:ða:* would be relatively more emphatic due to the presence of **ha:*. This opposition of *ha:ða:* to *ða:* would have allowed analogical extension of the **ha:*-prefix throughout the paradigm of the demonstratives.

Dialects which still oppose a demonstrative series with **ha:*-less forms and a series of forms with **ha:*- attached still show the difference in emphasis. W. Fischer (*ibid.*) notes that in these dialects, the forms without **ha:* act as ‘weak forms’ (Schwachform).

It is rare to find dialects with entirely optional **ha:*-prefixes throughout the paradigm. Usually part of the paradigm has obligatory **ha:*-prefixes, normally

⁶⁰ Original: “o Vater, da ist jenes Geld selbst”

on the singular demonstratives or in the entire near series. Dialects for which **ha:-*prefixes are apparently optional throughout the paradigm are ARBAH, ARANZ, AROM and possibly ARNJD. Dialects for which **ha:-*prefixes are obligatory only in the singular demonstratives include IQKHZ, AREM, LvSOU. Dialects where **ha:-*prefixes are obligatory only in the proximal demonstratives at present include LvPAL, UzBA, UzBDJ, though many North African dialects have **ha:-*optional distal adnominal demonstratives, but **ha:* is obligatory in the distal pronominal demonstratives, suggesting an earlier situation in which these dialects patterned with the **ha:-*obligatory proximal dialects.

The unbalanced nature of many of these paradigms, that is to say, with certain demonstratives requiring the use of a **ha:-*prefix, where in others it is option, suggest that first some forms grammaticalized with a **ha:-*prefix, and then the **ha:* was interpreted simply as a strengthening prefix. A simple analogy such as $\delta a::ha:-\delta a::?ula:X=ha:-?ula:$ would have spread the prefix throughout the paradigm. It also would seem that the origin of the change was in the near proximal demonstratives, based on the fact that these prefixes are either obligatory in all of the near demonstratives, or in all of the singular demonstratives. The association of **ha:-*prefixes with the near paradigm probably reflects the fact that the most common use of referent presentatives is to indicate something nearby or within the sphere of control of the speaker, and the higher frequency of these forms would lead to a more rapid grammaticalization. The same is true of singular demonstratives, which simply are likely to occur much more frequently than plural demonstratives, especially in Arabic where plural (especially non-human) collectives are often treated as feminine singular entities.

In many North African dialects, such as NATL, the **ha:-*prefix is obligatory for the pronominal demonstratives, but optional (or not present) for the adnominal

demonstratives, providing further evidence for the transition first from presentational or identificational demonstrative into a demonstrative pronoun, and only then into an adnominal demonstrative. In those dialects, it would appear that the process of attachment of the **ha:-* prefix was operative at the same time the demonstrative pronouns were grammaticalizing into reduced demonstrative determiners, but it **ha:-* had not yet spread through to the far demonstrative paradigm before the determiners were formed.

Though this scenario is somewhat complex, there is evidence for several of these stages. There is evidence for the transition between Stages 2 and 3, where *ha:ða:* was fused as a presentational demonstrative, but had not yet become a general demonstrative. This evidence comes from those dialects where the demonstratives do not contain **ha:* (and where the demonstratives are largely post-nominal), but which still have a sentence presentative use of **ha:ða:*.⁶¹ In (23), we see that this presentative form has lost enough of its demonstrative power that it is actually followed by a demonstrative which presents the sentence.⁶² See also the following example, where the $h > ? > \emptyset$, as often happens with the **ha:* form in Egyptian dialects (see *ibid.*, pp. 167-169).

(23) (SHUK, cited in *ibid.*, p. 177)

⁶¹ Many of these dialects have generalized a feminine version form of this presentative (W. Fischer, 1959, pp. 176-177), which is not entirely surprising given the tendency to generalize feminine forms. Indeed, in Yafi' Arabic discussed above and in (Vanhove, 2004) it is the feminine form *ta:* which has been generalized as a presentative and a copula. Moreover, W. Fischer (1959) shows other gender inflected forms which he considers later innovations, though they could just as easily be archaic retentions.

⁶² Note that SHUK is not truly a **ha:-* less dialect, since old speakers still optionally use near demonstratives with **ha:-*, though this might simply suggest that this dialect made it further in the process of **ha:* affixation and is now losing this to local sociolinguistic pressure, since this area is predominantly **ha:-*less. Example (24) shows an example from a more clearly **ha:-*less dialect.

wa-hadi 'da d'urri:jat faʃaddi:n
 and-PRESENT DEM descendents sha`addin

(and) these are the descendents of Sha`addin

(24) (Rural Egyptian, W. Fischer, 1959, p. 176)

a:di mi:ʃa:d ʃafa-lla fa:t
 PRESENT time fire-REL passed

Look, the time of the firefestival is already passed.⁶³

There is also evidence for Stage 4, where the presentative *ha:ða:* demonstratives, coming sentence initially, were reinterpreted as adnominal demonstratives. Most dialects without **ha:* affixed to the demonstratives show N-DEM word order, which was probably the original word order in Proto-West Semitic (Pat-El, 2009, p. 21), though as is typical for demonstratives, word order is often somewhat fluid (W. Fischer, 1959, pp. 64-65; Diessel, 1999, pp. 4-5, 61-2). However, dialects which have **ha:*-based demonstratives typically show DEM-N as the unmarked word-order, as noted by W. Fischer (1959, pp. 65-66).

This tendency of **ha:-* initial demonstratives to appear before the noun rather than after the noun in Arabic would suggest that these demonstratives are indeed derived from a sentence initial presentational demonstrative. A clear example of this difference between **ha:-*initial forms and **ha:-*less forms in terms of word order is seen in LvPAL, where Cantineau (1934, p. 220) notes that the **ha:-*less forms (“forme courte”) are used with N-DEM word order, while the **ha:-*prefixed forms are used with DEM-N word order. He gives examples *?etfe la ho:n ha:ðe:k*

⁶³ Original: “seht, jetzt ist die Zeit des Feierabends schon vorüber.”

ez-zalame (came to here that.M.SG DEF-man) ‘That man came here’ vs. *?etfe la ho:n ez-zalame ðe:k* (came to here DEF-man that.M.SG) ‘That man came here’.⁶⁴

This explanation does not account, however, for the tendency for demonstratives, even forms with **ha:-*attachment, to follow possessive clauses rather than proceeding them.⁶⁵ This, however, could be explained simply by the need to disambiguate between sentences which carry an identificational meaning, e.g. *ha:ði: bint-i* (this.F.SG daughter-mine) ‘This is my daughter’ and those in which the possessed nominal is the topic of the sentence (e.g. *bint-i ha:ði: qawiyy-a* (daughter-mine this-3SG.F strong-F) ‘This daughter of mine is strong’. Indeed, in many dialects the restriction is much looser, so that Cowell (1964, p. 558) gives the example *ha:da xayy-i ma: ha:zar* (DEM.M.SG brother-my NEG emigrate.PST.3M.SG) ‘This brother of mine didn’t emigrate.’

In contrast, with nouns defined using the definite article, the normal order in determiner phrase is DEM DEF-N, though the order can be reversed also. Here, though there is the ambiguity between the two interpretations of *ha:ða: ar-rajulu* as ‘this man...’ or ‘this is the man’, the phrasal rather than sentential meaning is much more conventionalized and more frequent. The sentential meaning is more marked, and likely to be indicated by context or pronunciation. In contrast, the sentential meaning of *ha:ði: bint-i* (i) ‘s’ likely more common, since the phrasal meaning (distinguishing between a number of different daughters) is going to be relatively rare. Thus, the conventionalized meaning of DEM DEF-N and that of DEM N-POSSESSIVE are different — the first is conventionalized to be understood as a phrase, while the second is conventionalized to be understood as a complete sen-

⁶⁴ He also notes that the long forms, i.e. those with a **ha:-*prefix are much more frequent, suggesting that the short forms are in the process of being replaced with the **ha:-*prefixed paradigm.

⁶⁵ They also tend to follow proper names, though this is a very marginal case.

tence. When a phrasal meaning is absolutely required, the demonstrative can be post-nominal in both cases; but this doesn't really change the conventional meaning of the DEM DEF-N construction — both pre-nominal and post-nominal positions have approximately the same meaning. In contrast, the N-POSSESSIVE DEM order is the only unambiguous conventionalized way to indicate that this is a phrase and not a complete sentence.⁶⁶

This account is supported by data from EGCAI where demonstratives are still primarily post-nominal, with pre-nominal demonstratives used for emphasis. In dialects with post-nominal demonstratives, there should be no ambiguity between a demonstrative phrase and a complete sentence with a demonstrative subject, since word order distinguishes between these two: DEM N will usually be a complete sentence (or rarely, an emphatic demonstrative phrase), while N DEM will almost always be a demonstrative phrase. Thus, we expect that demonstratives modifying genitive constructions would be in the same order as any other demonstrative construction, and this is indeed what we find, as shown in examples (25) and (26) where the demonstrative appears in its usual post-nominal position at the end of the genitive construction.

(25) (EGCAI, Woidich, 2006b, p. 198)

ʃaʔʔ-it-u di ka:n-it fi l-ʔabrʕa:g il-maʕa:di
apartment-GEN-his this.FEM was-3P.FEM in DEF-towers DEF-maadi

This apartment of his was in Maadi Towers.⁶⁷

⁶⁶ The difference is even clearer with proper names: it is much, much more likely that the speaker using a DEM PROPERNAME phrase is performing a presentative function, than distinguishing between people with the same proper name, and so the disambiguated PROPERNAME DEM order is much more common when a phrasal meaning is intended (see Cowell, 1964, pp. 557-8).

⁶⁷ Original: “diese seine Wohnung war in den Ma‘ādi-Türmen.”

(26) (EGCAI, Woidich, 2006b, p. 198)

ism il-binti di
name.GEN DEF-girl this.FEM

This girl's name.⁶⁸

This analysis of the development of *ha:ða:*-type demonstratives allows us to better understand a construction primarily found in Quranic and Classical Arabic that has typically been seen as related to the demonstratives. These constructions consist of a structure: [ha: PRO DEM], for example the final three words of example (27). The traditional analysis for these constructions is that they somehow represent the re-separation of the demonstrative into its component elements:

Zur Bezeichnung für das naheliegende 'dieser' verbindet sich *ða:* mit der Interjektion *ha:*; doch ist diese Verbindung noch so lose, dass sie durch dazwischentretende Pronn. (*ha: ?ana ða: [...]* *ha: huwa ða: [...]* *ha: hiya dihi*) und Parikeln (*ha: ?inna ða:*) getrennt werden kann.

For the meanings of proximal 'this', *ða:* is attached to *ha:*; however, this association is so loose that it can be split with an intervening pronoun (*ha: ?ana ða: [...]* *ha: huwa ða: [...]* *ha: hiya dihi*) or particle (*ha: ?inna ða:*). (Brockelmann, 1908, p. 318)

However, in the analysis here, these 'split' demonstrative constructions aren't 'split' at all, but rather syntactic traces of sentences in which a) **ha:* acted as a sentential presentative, and b) **ða:* was used in one of its normal functions within the sentence (either as an anaphore, or a sentential presentative). The explanation then is quite simple. In example (28) the function of the demonstrative is

⁶⁸ Original: "der Name dieses Mädchens."

anaphoric, that is it is referring to a previously mentioned discourse entity. The sentential presentative **ha:* is not strictly necessary: This sentence is entirely self-standing without it, but just like (29) we have some notion of presentation, and so we shouldn't be too surprised to find **ha:* used to introduce the sentence.

The function of **ǧa:* in (29) is different than its function in (28); instead of an anaphoric (endophoric) function, the demonstrative is functioning as a pronoun with exophoric reference to link that entity with the following information, i.e as a referential presentative. In this case the predicate is a pronoun, which necessarily has anaphoric reference, though the speaker could just as well say something like *ǧa: kanz-un* 'the Classical Arabic equivalent of (15) in which case there would be no anaphoric meaning per se. In contrast if the speakers had been speaking of a treasure, there would be no need to repeat the nominal, and so you'd have something like (29) meaning 'There it is [i.e. the treasure we were just talking about]'

(27) (Classical Arabic, from Bloch, 1986, p. 75)

qa:l-a qatal-tu Saduww-a lla:h-i bna xa:zim-in
 said-3m killed-1s enemy-ACC god-GEN.BOUND Ibn Khazim-GEN.UNBOUND
wa-ha: huwa ǧa:
 and-ha: he DEM.ms

He said, "I killed Ibn Khazim, the enemy of god, and here he is."

(28) Classical Arabic (posited example)

(ha:) huwa ǧa:
 (ha:) he DEM.ms

(Look,) he's that one.

(29) Classical Arabic (posited example)

(ha:) *ða:* *huwa*
 (ha:) DEM.ms he

(Look,) that's him.

This suggests that in structures like (28), the function of the **ða:-* demonstrative is essentially anaphoric, referring to an entity previously mentioned in the discourse, because that is the function in the sentence even without **ha:*. Indeed, the anaphoric use is exactly what we find in the Quranic examples. In (30), the demonstrative pronoun is clearly referring to those who 'dispute about Abraham', while the function of the plural second person pronoun *?antum* is to indicate who is indicated by this group. The anaphoric function of this pronoun is further reinforced by the poetic parallelism of the verses, which use the same verb (*ħa:jj* 'to dispute') in both the first and the second verse. Other verses showing using this construction, such as verses 3:119 and 4:109 also show almost identical parallelism, including the use of the same verb in both the initial verse, and the following verse with the HA: PRO DEM construction.

(30) Quran 3:64-65

يَا أَهْلَ الْكِتَابِ لِمَ تُحَاجُّونَ فِي إِبْرَاهِيمَ وَمَا أُنْزِلَتِ التَّوْرَةُ وَالْإِنْجِيلُ إِلَّا مِنْ بَعْدِهِ أَفَلَا تَعْقِلُونَ (٦٥) هَآأَنْتُمْ هَآؤَآءِ حَآجَجْتُمْ فِيمَا لَكُمْ بِهِ عِلْمٌ

ya: ?ahl-a l-kita:b-i lima tuha:jjun fi: ?ibrahi:m ma: ?anzalt-i
 oh people of-the-book why you-dispute about Abraham NEG revealed
t-awra:t-u wa-l-?inji:l ?illa: min ba?dihi ?afala: ta?qilu:n. ha ?antum
 torah gospel only since after-him do-NEG you-think. ha: you.pl
ha:?ula:?i ha:jajtum fi:ma lakum bihi.
 DEM.near.pl disputed-you.pl in-what to-you.pl in-it knowledge

(Verse 64): O people of the Scripture (Jews and Christians)! Why do you dispute about Ibrâhim (Abraham), while the Taurât (Torah) and the Injeel

(Gospel) were not revealed till after him? Have you then no sense? (65)
 Verily, you are those who have disputed about that of which you have
 knowledge...

Moreover, this kind of construction should really just be seen as a subset of the much more common set of sentences of the type HA: PRONOUN PREDICATE which are so common that we often find dialects where **ha:* is no longer self-standing, but must have a suffixed pronoun, which is often followed by further information, but where the sentence could be self-standing without **ha:* (W. Fischer, 1959, dedicates an entire section to these, see pp. 164-166). An excellent example of this is (13), while a HA: PRONOUN sentence is exemplified in (14).

The very same can be said of the constructions which Brocklemann notes where **ha:* precedes a sentence with the emphatic particle *?inna*. W. Fischer (2002, §279) gives (without attribution) an example from the grammarian al-Siyūṭī (d.1505) show in (31). There is no special meaning for the demonstrative here – it simply functions as an anaphore, and occupies its normal subject position in the sentence. This type of structure is vanishingly rare in actual usage, as far as I can ascertain from a search in the arabiCorpus’s somewhat limited Classical Arabic corpus. However, a structure with *ha: ?inna...* with no demonstrative in the following phrase is quite common, and this should simply be seen as a subset.⁶⁹

(31) Classical Arabic

⁶⁹ I only found one use with a demonstrative (outside of a grammatical treatise) in arabiCorpus’s pre-modern corpus is from Kitāb al-Aghani, *ها إن ذا لا يقر الطير ساكنة ولا تبرك من نكرائه* *الإبل* ‘Look, that certainly does not make the bird sit still and quiet, or make the camel kneel down from its disobedience.’

ha: ?inna ði: ?uðrat-un ?in la: ta-kun
 ha: indeed DEM.F.SG.PROX excuse-NOM if NEG 3f.sg-to.be
nafaʕ-at
 to.be.useful.PSF-3f.sg

ha:, now there's an excuse, even if it does no good.

The one surprising aspect of this example is the use of *ði:* instead of *ha:ðihi*; Here, speakers (in this case, a grammarian) educated in a tradition which emphasizes the importance of structures such as the (28) might simply have generalized the structure further to this type of structure, or the examples may be from a time before **ha:* and **ða:* grammaticalized into a single element.

It is also worth noting the difference between this structure and a very similar set of structures in Biblical Hebrew. In Biblical Hebrew, the conjunction of **ha:* and *?inna* were apparently much more frequent, and thus *hinne* < **ha-?inna* appears to have grammaticalized as a presentative, much as **ha:-*ða:* did in Nile Valley dialects. Moreover, the structure *hinne ze* (PRESENTATIVE DEM) had, for the masculine at least, formed a grammatical entity so that the demonstrative no longer has any place in the following sentence. This is not true for feminine or plural demonstratives, as can be seen in the following contrasting examples. The (originally) demonstrative form *ze* has no syntactic role in (32), but in (33) the demonstrative *zō't* occupies the position of subject in the sentence.

(32) (Biblical Hebrew, 1Kg 19:5, Huehnergard and Pat-El, 2007, p. 336)

w hinne ze mal'āk nōgea' b-o
 and PRESENTATIVE DEM.(REL).MS angel.MS touch.PARTICIPLE.MS in-him

Then an angel touched him

(33) (Biblical Hebrew, Ezek 19:5, *ibid.*, p. 336)

<i>hinne</i>	<i>zōʾt</i>	<i>torat</i>	<i>hab-bayit</i>
PRESENTATIVE-DEM.FS	law.FS.BOUND	DEF-house.MS	

This then is the law of the house.

This had led Huehnergard and Pat-El (2007) to argue that in these masculine presentative constructions in Hebrew, the demonstrative acts as a type of relative marker, especially since the form in question could be used as a relative marker elsewhere. However, Arabic has clearly not followed the same path as Hebrew – in both example (31) and the example in footnote 69, the demonstrative is clearly occupying the subject position, and therefore it is much simpler to analyze this sentence as the base sentence beginning with *ʔinna* ... which happens to have a **ha*: presentative preceding it. That is to say, the Arabic examples regardless of gender and number look more like (33) than (32).

5.4.3 **ha*: as a demonstrative

The **ha*: presentative has also developed into two important demonstratives, one of which is as a demonstrative determiner, used to mark nominals which are psychologically close-at-hand, and as a full demonstrative pronoun, which has both pronominal and adnominal functions. The most widespread of these two is the use as an demonstrative determiner, with no original marking for distance, gender or number, and it is probably historically the first to develop. Most dialects which have **ha*:- as a demonstrative determiner do not have it as an independent pronominal demonstrative, suggesting that the development of the pronominal demonstratives are a later innovation.

5.4.3.1 **ha*: as an adnominal demonstrative

The development of the adnominal use of this demonstrative probably comes from contexts such as (34) or (11), where the sentential presentative, preceding a sentence with an initial definite could be interpreted as having the same syntax as the demonstrative constructions with DEM DEF-N⁷⁰

(34) (Shloh Arabic, Morocco: W. Fischer, 1959, p. 158)

waħəd ən-nħa:r^s ħa: d-di:b ʒai ka:ma l-ʕa:da dia:l-u
 one DEF-day, ħa: DEF-wolf coming as DEF-custom GEN-his

One day, look, the wolf is coming, as is his custom.⁷¹

This demonstrative determiner is only adnominal and does not make a distance contrast, and therefore one of the few functions it can perform is a recognitional function, as discussed in Section 5.1. However, the definition of a recognitional demonstrative according to Diessel (1999) is to introduce discourse-new, hearer-old information into the discourse, a function that the definite article can also fill. How then does the **ha*: demonstrative differ?

It seems also that the primary restriction on **ha*: is that it is used to mark entities which are psychologically close-at-hand, that is to say they are easily retrievable from the discourse context, or from the nearby discourse itself (i.e. previous utterances). In Tunisian Arabic, the **ha*: demonstrative can only be used

⁷⁰ This order would have existed even in dialects with primarily DEF-N DEM word order. Indeed the fact that this order is typically more emphatic might have helped with the interpretation of **ha*: not just as something emphasizing the sentence, but instead as a demonstrative determiner.

⁷¹ This is translated as an expression of surprise in the original, “eines Tages, da kam plötzlich der Wolf, wie es eine Gewohnheit war” ‘One day, a wolf suddenly came, as was his custom’. This expression of surprise seems to contradict the following clause, and so I prefer a reading where the element *ħa*: is operating to shift the scene to the one in which the wolf appears, instead of as a marker of some sort of surprising event. However, the function of calling attention would of course be associated with surprising circumstances.

with referents which at least have the status of activated in the discourse (Khalifaoui, 2007), while the definite article can mark anything from entities which are in focus to entities which are simply type-identifiable. Khalifaoui (2013) showed that the **ha:* determiner in Tunisian frequently occurs with the second mention of newly introduced characters in folk narratives, a use which suggests a certain recentness is required for the use of this determiner. Similarly, Brustad (2000, Chapter 4) argues that **ha:* determiners mark only identifiable, textually prominent referents, whether they have been previously activated or are schematically activated by some aspect of the discourse.

These restrictions make sense from the perspective of the historical development of **ha:* determiners from presentative, exophoric demonstratives. They also explain the use of **ha:* determiners to shift the deictic center from the real world to the narrative, called by Noonan (2001, p. 179) NARRATIVE DISPLACEMENT⁷², “a shift of the actual situation of the utterance to the temporal and physical setting of the narrative itself, wherein the speaker pretends that the narrated events are happening right in front of the speaker and audience.” The mechanism that causes this effect is the restricted ‘range’ of this demonstrative — if the speaker refers to a place that does not exist, the listener must conjure that place for the demonstrative use to make sense. This is very clear in examples (34) and (35), especially the latter where the definite article would have had a significantly less dramatic effect. Many more examples of this can be seen in W. Fischer (1959, pp. 158-159).⁷³

(35) (Syrian Arabic: Brustad, 2000, p. 123)

They would go every Saturday for a whole year to light incense in the early morning...

⁷² This is also called by Bühler (1934, p. 121) “Deixis am Phantasma”, see Himmelmann (1996, p. 222)

⁷³ The same is true of similar demonstratives derived from **ha:* in Biblical Hebrew and Aramaic.

bi-ru:h-u b-ha l-fatme w b-ha l-le:l yru:h-u yfaʃʃlu baxxu:r
 IND-go-PL in-ha: DEF-dark and in-ha: DEF-night go-PL light-PL incense

They go in that darkness and in that night to light incense ...

This transition from a presentative, and often sentential presentative to a recognitional, adnominal demonstrative also explains the co-occurrence of **ha:* with formally definite nouns, particularly with the definite article. This goes against the expectation for a recognitional phrase which introduces something new into the discourse — a function often carried out using an indefinite. However, nominal sentences in Arabic almost always begin with a formally definite noun. The presentative particle was sentence initial, an artifact of its original function as an attention getting particle, and so would have been in loose juxtaposition with the following sentence, which is then reanalyzed as a DEM-N initial sentence. For this reinterpretation to happen, however, that initial nominal would almost certainly have to be definite, as it is in both (34) and in (11) where the **ha:* precedes a noun made definite with a pronominal suffix. Thus, in most dialects, the recognitional adnominal demonstrative derived from **ha:* is generalized with a following definite article, producing strange constructions as in (36) where the **ha:* demonstrative is strangely transferred to the adjective in what appears to be a type of agreement phenomenon.

(36) (Syrian Arabic: Brustad, 2000, p. 122)

halla? ha-j-ji:l ha-l-mawju:d b-uxtʃb-u baʃdʃon
 now ha:-DEF-generation ha:-DEF-present IND-get.engaged-PL each.other
hinnin
 they

This generation they arrange their own marriages

In some dialects, such as the ancestors of many north African dialects, **ha:* probably did not undergo this reanalysis as an adnominal demonstrative, where a slowly grammaticalizing demonstrative set with *hadV* was reduced to a gender neutral *had-*, and which developed a recognitional use simply as one of the many functions that an adnominal demonstrative carries. Here the still purely presentational **ha:* particle can precede a valid sentence, as in (37), where it precedes a noun phrase marked by an indefinite article. This may be related to the fact that these North African dialects innovated indefinite articles which were much more widely applicable than in other dialects (Brustad, 2000, 31-36ff), so that **ha:* would often precede nouns without a definite article, as in (37). In those dialects, the reanalysis would therefore not be triggered, and **ha:* would not become a generally demonstrative determiner, but remained in its presentative uses.

(37) (Mogador (Essouira), Morocco: W. Fischer, 1959, p. 158)

fwoija fwoija ha wa:hid fəxs^s da:xil
 little little ha: ART.INDEF person entered

After a short time, look, a person entered⁷⁴

5.4.3.2 **ha:* as a pronominal demonstrative

It would have also been easy to reinterpret **ha:* as a pronominal demonstrative, since it often functioned as a referent presentative, which in Arabic looks very much like a subject in a copular sentence (since Arabic does not require an explicit copula.)⁷⁵ In example (8), the **ha:* used in this sentence is uninterpretable

⁷⁴ This is also translated with “plötzlich” in Fischer’s text, an expression of surprise.

⁷⁵ It is interesting that the use of **ha:* as an adnominal demonstrative is much more widespread than the use of **ha:* as a pronominal demonstrative. Given the possible bridging contexts, both uses

without the exophoric context, but it is also a potential bridging context in which **ha:* could be interpreted not as an interjection but as part of a equational nominal sentence. Such contexts would allow for a development of **ha:* into a demonstrative pronoun, which may have already happened in this dialect.⁷⁶

It is unclear whether the pronominal **ha:* demonstratives developed directly from the referent presentative use, or from the adnominal use. It seems likely that it might have been both contexts together which allowed for the interpretation of **ha:* as a full-fledged demonstrative, and not just a presentative. This would be supported by the spread of **ha:* — its use as an adnominal is much more wide-spread than its use as a pronominal demonstrative, suggesting that its frequent occurrence in contexts as a referential presentative was not sufficient to trigger reinterpretation directly. Once it had become an adnominal demonstrative though, it would have been easier to make that reinterpretation.

In the dialects which did reinterpret **ha:* as a demonstrative pronoun, it can act in largely the same ways as the normal, (**ha:*) **ǝa:* series of demonstratives: as a pronoun (38), as an adnominal demonstrative (39) (in this dialect demonstratives are often post-posed, but see Section 5.4.2.1), as a referent presentative (40) or as a sentential presentative (41).

(38) (Syrian Arabic, Cowell, 1964, p. 554)

kəll fi: ʔəlla hayy walʔla
any thing but ha:F by.god

are equally likely to develop. One wonders whether there might not have been some influence from ANA, which had an pre-nominal article *ha(n)* — contact with such languages might have tipped the scales in favor of interpreting **ha:* as an demonstrative determiner.

⁷⁶ It would be interesting to know if, in the Kuwaiti dialect, whether this sentence could have been uttered with an explicit copula, *ha: huwa tu:l-a* (ha: COP length-its). If so, then **ha:* can clearly function as a pronominal demonstrative. If not, then it is still an interjection.

Anything but that!

- (39) (Palmyra, Syria, Cantineau, 1934, p. 220)

mne:n al-ħor^ʕm^ʕa he?
from.where DEF-woman ha:.C

Where is that woman from?

- (40) (Syrian Arabic, Cowell, 1964, p. 553)

hayy al-bənt yalli ʔal-t-əll-ak ʕan-ha
ha:.F DEF-girl REL told-1s-to-you.M about-her

Here's the girl I told you about

- (41) (Syrian Arabic, *ibid.*, p. 554)

ya ʔax-i l-ʕazi:z, hayy ʔante yalt^ʕa:n ta:ni marra
VOC brother-my DEF-dear, ha:.F mistaken another time

My dear friend, there you're wrong again.

5.4.4 Elaborated **ha:* forms

At some point in their development, **ha:* demonstratives also developed gender, and in some cases, plural marking, as is clear from the examples above. What is unclear is when in the **ha:* demonstratives, first began to be marked for gender and number: when they were still only functioning as adnominal demonstratives, or once the pronominal **ha:* demonstratives had developed.

The addition of gender and number marking suffixes could have occurred even while **ha:* was operating only as a demonstrative determiner. In many dialects (especially Tunisian dialects, but including many Mesopotamian and Levantine dialects), **ha:* has developed a distal counterpart, *ha:k*, simply by adding the

distal marker *-k*. This distal counterpart is still restricted to an adnominal function. A similar type of development could have lead to the creation of adnominal feminine *hay* adding the apparent formative *-i* based on the connections illustrated in Table 5.6 to *ha:*, as well as the development of *haw* via the addition of the apparent formative *u*.

What is clear is that the feminine demonstrative **ha:-i* is much more successful and widespread than its masculine counterpart *ha:*. The **ha:-i* is widespread in Mesopotamian dialects as *ha:ʔi* (W. Fischer, 1959, pp. 52-53), where a glottal stop is inserted between adjacent vowels to avoid a diphthong, as in *abu:ʔi* < *abu:-i* (father-mine) ‘my father’ or *fi:ʔu* < *fi:-u* (in-it) ‘in it’.⁷⁷ In the Levant, forms with a glide are common, such as *ha:y* and *hayy* where the length of the vowel has been apparently replaced by gemination of the glide.

In many of the dialects with this feminine form, there is no masculine pronoun from the same source, and instead the form *ha:ða* is retained, with *ha:y* functioning as the primary proximal feminine demonstrative (many of these dialects retain *ha:ða:k* in the distal forms). Moreover, in dialects which did have an original *ha:-*type pronominal demonstrative such as LvSOU (where the form is gender neutral, just like the adnominal equivalent), the diffusion of *hay* between population centers in the Levant means that the *hay* demonstrative has actually been borrowed in that dialect (Behnstedt, 1994, p. 116).

It would appear at first glance that some dialects have innovated *haw*: for example, for LvBSH: Jiha (1964) gives *hawdi*, *hawd*, *haw* as demonstrative pronouns, and the Anatolian dialect of Mardin has a distal demonstrative (pronoun

⁷⁷ Obviously from the first example, this isn’t necessarily a change of *h* > *ʔ* though it could potentially have started on the third person pronoun and moved to diphthongs more generally. However, Syriac also uses a glottal stop to break up diphthongs, so this is not entirely unusual.

and adnominal) *hawk*, though there does not at present appear to be a proximal *haw*. However, once we look closer we noticed that, in actual usage *haw* seems to be almost never found as a demonstrative pronoun that is clearly derived from *ha:* directly. In LvBSH the reduced forms tend to come only in utterance-final position, so these are clearly derived from the longer form. The same seems to be true of other dialects with these forms – Behnstedt (1997, map 277) shows a small cluster of dialects just north-west of Idlib (as well as LvBSH of course) which have *haw(we)* forms, but they also show *hawdi*, again suggesting that *haw* demonstrative pronouns are actually derived from a shortening process. The actual development of *haw* is much more complex than it seems, and is the subject of the following section.

5.4.4.1 Elaborated **ha:* forms with **ǵa:*

There are, especially in the Levant and particularly in what is now Lebanon, demonstratives which have an elaborated **ha:-*type prefix on their **ha:- *ǵa:-*type demonstratives. In Lebanon, a feminine proximal *haydi* is widespread, opposed typically to masculine *ha:da*, and almost always with plurals like *hawdi*, again much like LvBSH with *hæ:da*, *haydi*, *hawdi*. Similarly, in Anatolian Arabic dialects we find a number of dialects which have plural proximal demonstratives derived from either **hawǵa* or **hawǵi*, though only a subset of these dialects have feminine singular proximal **hayǵi* (Jastrow, 1978, pp. 102-108).

Indeed, it is rather mysterious that in these dialects, many of which have a pronominal demonstrative *hayy*, the *hawǵV* forms appear to be historically prior. If this was simply reinterpretation of *hayy* as equivalent to the *ha:-* prefix on verbs, we would expect the opposite, especially given the rarity of free *haw* demonstrative forms. There are numerous dialects in both the Levant and Anatolia which

show those forms but which do not have *hayði*-type forms, as illustrates by Table 5.10.

Table 5.10: *hawd*-type demonstrative paradigms

	Syria #284 (Just SE of Idlib)		Diyarbakır (Anatolia)	
	Proximal	Distal	Proximal	Distal
M.SG	ha:da	haða:k	(h)a:d(a)	(h)a:k
F.SG	ha:y	haði:k	(h)a:di	(h)ayk
C.PL	hawdi	hawði:k~hawki	(h)awd(e)	(h)awk

Both of the dialects illustrated in this chart have *hawd*-demonstratives, but no *hayd*-demonstratives. Where then does this *haw*-prefix come from? Certainly, it seems reasonable to develop a **ha-u* > *haw*-type demonstrative in analogy with *ha:y*. But note that there is a difference: **ha:-i* often retain the length of the *ha:*, as in Syria #284's *ha:y* or the syllable is restructured as *hayy*. This is one clue that suggests a different source for *haw*, and indeed it seems to come not from extending *ha:*, but instead from a back-formation of another, entirely unrelated demonstrative, the plural form *hawl(V)*, originally from *ha:-ʔula* (PRESENTATIVE-DEM.PL).

It is necessary, of course, for speakers to have access to that demonstrative *hawl(V)* in order to make this backformation, which is clearly absent from the dialects of Syria #284 and Diyarbakır. However, the plural demonstrative forms in these dialects are clearly innovative, and there must have been a different plural demonstrative in their ancestors. One can see a relic form in Syria #284's *hawki* which is clearly not derivable from the other forms, and which is not found in surrounding dialects. The original form of the plural demonstratives in these dialects was, it would seem, *hawl(V)*. That demonstrative is still found in Anatolia in the

dialects of Siirt (see Table 5.11), as well as scattered throughout the regions of the Levant which show *hawd*-type dialects.

Table 5.11: Anatolian dialects with *ha:k* demonstratives

	Siirt		Aghde (Sason-Type) Dialect	
	Proximal	Distal	Proximal	Distal
M.SG	a:va	ɑ:k	a:za	a:ga
F.SG	a:vi	ɑ:k	a:zi	a:gi
C.PL	awl(e)	awlok	a:zu	a:gu

The scenario for the development of *hawd*- demonstratives is as follows, and is illustrated in Table 5.12: In these dialects, the near, general adnominal demonstrative *ha:k* (possibly originally *ha:ka*) is innovated from *ha:* using the transparent *k(a)* distal marker.⁷⁸ This marker, as it currently is in Tunisian dialects, would have marked all genders and numbers.⁷⁹ In a process similar to that which produced *ha:y*, this adnominal would have become a general adnominal and pronominal demonstrative which could act as a pronoun for referents of all number and genders. It would have therefore replaced all of the distal demonstrative form (Stage 1)

While there is not a great need to mark gender or number in an adnominal demonstrative (it is marked on the nominal it attaches to), it is much more

⁷⁸ This means then that these dialect originally must have had distal demonstratives that were clearly derived from the proximal demonstratives with this *-k(a)* distal marker, something like *ha:ða*, *ha:ða:k* and *hawla*, *hawla:k*.

⁷⁹ W. Fischer (1959, pp. 45-6) argues that *ha:* and hence *ha:k* in North African dialects which have this feature are actually derived from the erosion of *ha:da* > *ha:d-* > *ha:* so that they are not actually from the same source originally as the Levantine and Iraqi *ha:k* demonstratives, but rather a process of parallel evolution. It is also possible that Tunisian retained or evolved *ha:* as a demonstrative in the same way that it evolved in other dialects, from the well attested use of *ha:* as a presentative in North African dialects.

important for a pronominal demonstrative, where there might be ambiguity in what referent it indexes. In some dialects, such as that of Aghde (see Table 5.11), where the proximate demonstratives were **ha:da*, **ha:di*, **ha:du*, the solution is quite simple, since the endings are easily analyzable, so we get **ha:ka ha:ki ha:ku* (with subsequent changes in the consonants and loss of /h/).

For dialects whose ancestral dialect ('Stage 0') was something like that in Table 5.12, however, the solution is not so neat.⁸⁰ However, there is the demonstrative *hawl(a)* in the proximal, which looks very similar to *hak(a)*. It is a simple logical leap to interpreting *hawl(a)* which at this point is probably not analyzable back into its original *ha-?ula*, as something like *haw-la*, equivalent to *ha-ka*, creating the analogous form *hawka* (Stage 2.)

In Stage 3, the proximal plural demonstrative does not fit with the rest of its paradigm, in contrast to the distal, and is then levelled, though it appears that dialects chose differently as to whether to level based on the masculine or the feminine form (for discussion, see Jastrow, 1978, p. 104).

In Stage 4, the feminine singular distal demonstrative is extended to match *hawk(a)* by analogy with *ha:y*, a stage which is necessary to account for dialects like Diyarbakır (Table 5.10) where the feminine proximal is still *(h)a:di*. Finally, in Stage 5, the proximal feminine is finally adjusted to *hayđi* in analogy with *hawđ(a/i)*.

There are no dialects to my knowledge which still show the development in Stage 1, though the Sason dialects in Anatolia are close, which have evolved gender markers at the end of the *ha:k* demonstratives, so they have **ha:ka*, **ha:ki*, **ha:ku*. Stage 2 dialects are also extremely scarce, but this stage may have set up an

⁸⁰ The development and geographical extent of these dialects is discussed in Section 6.2.

Table 5.12: Stages in the development of *hawd*- dialects

	Stage 0		Stage 1		Stage 2	
	<i>Proximal</i>	<i>Distal</i>	<i>Proximal</i>	<i>Distal</i>	<i>Proximal</i>	<i>Distal</i>
M.SG	ha:ða	ha:ða:k	ha:ða	ha:k(a)	ha:ða	ha:k(a)
F.SG	ha:ði	ha:ði:k	ha:ði	ha:k(a)	ha:ði	ha:k(a)
C.PL	hawl(a)	hawla:k	hawl(a)	ha:k(a)	hawl(a)	hawk(a)
	Stage 3		Stage 4		Stage 5	
	<i>Proximal</i>	<i>Distal</i>	<i>Proximal</i>	<i>Distal</i>	<i>Proximal</i>	<i>Distal</i>
M.SG	ha:ða	ha:k(a)	ha:ða	ha:k(a)	ha:ða	ha:k(a)
F.SG	ha:ði (ha:y)	ha:k(a)	ha:ði (ha:y)	hayk(a)	hayði	hayk(a)
C.PL	hawð(a/i)	hawk(a)	hawð(a/i)	hawk(a)	hawð(a/i)	hawk(a)

association that had much larger consequences for the dialectology of Syria and which suggests that these dialects were widespread at one time. The difference between proximal plural *hawl*- and *hawk*- would have suggested to speakers that some sort of formative /l/ marks proximal status in plurals, while /k/ marks distal status. This incorrect analysis explains a series of forms that are quite wide-spread in the area around Homs, Syria, where the proximal plural forms are **haðo:l* but the distal plural forms are **haðo:k*. As I will argue in the following chapter, the original dialects in this area were likely of the *ha:k* variety discussed here, but later a set of dialects which had a proximal plural form *haðo:l*, distal *haðo:la:k* moved into the major urban areas, and spread between the cities. The interaction between these two dialects, especially if the *ha:k*-type dialects were at Stage 2 at this time, would explain these unusual forms, instead of falling back on some process that deleted the syllable /la:/ in the distal *haðo:la:k*.

Stage 3 dialects are extremely widespread and well attested, especially in the mountains of NW Syria, but also in the Anatolian dialect of Mardin. Stage 4

as mentioned earlier is attested in Diyarbakır. Stage 5 dialects are attested, but as in Beiruti Arabic, the distals are often reformed based on the singular in another stage of development not featured here.

Another stage not included in the diagram created a form *hayda*, as in Beiruti Arabic. The reason for this development may be essentially phonological. There are many Lebanese dialects where the masculine proximal demonstrative has a raised vowel in the first syllable, e.g. *he:da*. This raised vowel /e:/ is typically the reflex of original /*ay/, but in almost all of these dialects, /*a:/ has also been raised to a mid-vowel /e:/. Thus, there is a merger that means that the original vowel is not necessarily clear: /e:/ could either reflect original /a:/ or /ay/. There are, however, some diagnostic dialects, which realize original /*ay/ as /ey/ instead, and /*a:/ as /e:/, such as Syria #327 which has the demonstrative *he:de*. Thus, that demonstrative form is almost certainly from **ha:da* and not from an original *hayda*. Nonetheless, this ambiguity may be the impetus for reinterpreting forms like *he:da* as originating in *hayda* and adjusting the demonstrative accordingly, along with the paradigmatic pressure of *haydi*.⁸¹

Indeed, Naïm (*EALL : “Beirut Arabic ”*) gives a form *haydu:l* for the plural, probably originally **he:do:l* < **ha:ðo:l*. The phonological similarity to *hay-* would have allowed for the complete levelling of the *hay-* prefix throughout this paradigm, in spite of its original association with the feminine.

⁸¹ Of course, the innovation that created *hayda* seems to post-date the shift of original /ay/ > /e:/, which is why it is realized as a diphthong, but speakers of many of these dialects would have been in contact with numerous exemplars from other dialects (especially in cities such as Beirut) that would have allowed this analysis, as well as potentially exemplars from Standard Arabic where the diphthongs are prescriptively realized.

5.4.5 Word-final *-ha:*

In a number of dialects, particularly in the Nile Valley and areas south, the (especially distal) demonstratives have final suffixes of the form *ha:*, almost always attached to demonstratives of the form *dVʔk*. For example, EGCAI has the demonstrative paradigm shown in Table 5.13, with the very rarely used distal demonstratives included.

Table 5.13: EGCAI Demonstratives

	m. sg.	f. sg.	m. pl
Proximal	da:	di:	do:l
Distal	dukha	dikha	dukham~dukhum~dukhumma

It is not immediately clear whether these suffixes on the distal forms represent some sort of affixed pronoun, or whether the *-ha* in the singular forms represents a post-posed version of the **ha:-* interjection. W. Fischer (1959, p. 95) very clearly takes the latter view – in presenting his section on suffixed *-ha:* he begins:

Die deiktische Interjektion *hā*, die gewöhnlich den demonstrativen Formen präfigiert wird, findet sich in manchen Dialekten auch als Suffix.

The deictic interjection *ha:* which normally is preposed to the demonstrative, is also found in many dialects as a suffix.

In this view, then, the apparent similarity between the feminine pronominal suffix *-ha* and these demonstrative forms with final *-ha* is coincidental, and

would be strengthened by the fact that in the masculine form, it is still *-ha* and not the masculine pronominal suffix *-u*. For the similarity between the plural pronoun suffix *-hum* and the forms *dukham(ma)*, W. Fischer (1959, p. 96) argues that the vowels here reflects an original form **dikha:-humma*, with either convergence to /a/ or /u/ producing either of the attested *dukhum* or *dukham* forms.

W. Fischer (*ibid.*) does not however explain how **ha:* came to be attached to the end of the demonstrative forms, considering its origins as a sentence and clause initial particle.⁸² Indeed, there does not seem to be any clear functional motivation for such a change. Instead, I suggest that this final form is simply the feminine pronominal suffix used without an endophoric anaphoric referent within the discourse, a usage which is found in Egyptian Arabic. (Woidich, 2006b, p. 356) describes several such uses of *-ha* in Egyptian Arabic: As an anaphoric discourse demonstrative, referring to a previous general idea rather than a single referent, as in (42); Used as a reference to spatial or temporal distance, or objects outside the discourse, as in (43); or referring tacitly to a taboo idea.

(42) (EGCAI *ibid.*, p. 356)

ʕirif-t-aha mitʔaxxarʕ ʔawi
 knew-1s-ha late very
 I learned of it quite late⁸³

(43) (EGCAI *ibid.*, p. 356)

hasan mifi:-ha ha:fi
 hasan walked-ha barefoot
 Hasan walked [this street] barefoot.⁸⁴

⁸² One is also struck by the opposite patterning between these forms — almost all distal — and the **ha:*-prefixing, which generally appears to start in the proximal demonstratives.

⁸³ Original: “ich habe es sehr spät erfahren.”

⁸⁴ Original: “Ḥasan lief [diese Strecke] barfuss”

This would explain, functionally, the reason for the similarity between the masculine and feminine demonstrative forms. The pronoun is not referring to an indicated object per se, but rather is a general pronoun-demonstrative, derived from the feminine singular pronoun suffix.⁸⁵ In the masculine and feminine, this singular form is sufficient, though it would appear that speakers have by analogy (or another process, see below) developed plural forms based on the plural pronouns. The /a/ in these pronouns (i.e. *dukham* instead of expected *dukhum*) may be some sort of attachment of the type proposed by Fischer above, or may simply be contamination from the other members of the paradigm, which in both the masculine and the feminine end in a final /-a/.

Moreover, W. Fischer (1959, pp. 96-7) himself notes that these demonstrative-final forms look like pronominal forms, since in a number of dialects which have similar far demonstrative forms, there are /t/-final forms, e.g. *dikhat* (unspecified Egyptian form). As Fischer notes, many Arabic dialects show /t/-final pronominal forms, inside and outside of Egypt, so within Egypt we have forms such as Cairene *ʔa-huwwat* (PRESENTATIVE-3m.sg) ‘There he is’, or even Mauritanian *hu-ma:ti* ‘they’, and /t/-final demonstrative forms are found throughout the Semitic languages, e.g. OSA HWT ‘he’, Guarage *xuta* ‘he’ and in the Afro-Asiatic Egyptian language. Behnstedt and Woidich (1985, map 148) show that large swathes of Egypt have forms such as *huwwat* ‘he’, *diyyat* ‘feminine proximal demonstrative’, especially in the Delta. The *-ha:* at the end of these demonstratives then is clearly

⁸⁵ In this capacity, it may well be related to the original function of the forms which became third person pronominal suffixes in Semitic, but which Huehnergard and Pat-El (2012) argue were originally demonstratives. However, it is not clear whether this use is truly demonstrative, which would mean it is directly related to the archaic demonstratives, or is just the use of a pronoun for contextual use, a function which is often carried out in Semitic languages using the feminine (cf. their arguments on p.37). The contextual interpretation seems more felicitous here, since these demonstrative-presentative forms are clearly pointing to something in the discourse context.

derived from a pronoun, and not from the presentative **ha:*.

It is also unusual for Arabic dialects to have pronouns affixed to demonstratives that mark the object being pointing to, and not the interlocutor. Quranic Arabic, ARQAH and the Arabic of Jabal Razih in Yemen have all reinterpreted final *-k* on their demonstratives as the second person pronoun, so they produce forms like *ḏakum* (cf. the plural second person pronoun suffix *-kum*) is used when speaking to a group of males (Watson et al., 2006a; see also Michalski, 2011, regarding Quranic, Classical and Modern Standard usages of this structure). This development is clearly an innovation in a small number of Arabic dialects, due to the apparent similarity between distal-marking *-k* and the pronoun suffixes *-k(a/i/um)* (Rebecca Hasselbach, p.c.) and does not occur elsewhere in Semitic, or even very commonly in Arabic, which in some ways is remarkable given the relative ease of such a reinterpretation. It is also fairly uncommon cross-linguistically — Dixon (2003, p. 67) notes that some languages affix third person pronouns to their demonstratives, but he has no examples with second person pronouns.

Where Arabic and Semitic languages do have pronoun suffixes on its demonstratives, they are normally part of some sort of presentative structure, where a referent presentative acts as the subject of a clause, and a (independent) pronoun acts as a predicate, though in some dialects the structure is altered to use a suffix pronoun instead, though the presentative+pronoun structure is still interpretable as a full clause.⁸⁶ So in CHDSH, they have presentative structures like *da:hu:* ‘this is him, here he is’ and *do:la huma* ‘this is them, here they are’, but also a fused form *dowa* < **da: hu:wa* meaning *here it is, look* (W. Fischer, 1959, p. 180). In EGCAI, the structure is clearly derived from independent pronouns, i.e. *da-na*, *da-nta*, as

⁸⁶ For example, Hebrew *zehū* ‘this is him’.

is their presentative derived from **ha:* e.g. *a:hu:wa* (< **ha hu:wa*), but there are some ambiguous forms such as *a:hu:* which bears a greater resemblance to the third person masculine pronoun suffix *-u*. Similarly, in other North African dialects, we find clear uses of the pronoun suffixes, as in the dialect of al-Ḥamma in Tunisia, we have *ha:k* ‘here you are’ *a:ni* < **ha:ni* ‘here I am’ where the endings are more clearly derived from suffix pronouns (W. Fischer, 1959, p. 168).

In fact, in Nigerian Arabic (NIG), the forms which are cognate to those in Egyptian are used as presentives. In Nigerian Arabic, a representative sample of the demonstratives and presentatives is show in Table 5.14.⁸⁷ Note the presentative forms very clearly had an affixed personal pronoun marked for gender (masculine *huwa*, feminine *hiya*), in addition to the marking of the base with regard to the same gender and number. Indeed, the only other demonstrative-type forms which have pronominal suffixes in Arabic are those forms which operate as presentatives, such as EGCAI *?a-ho* ‘There he is’, NACHR *ha-hu* ‘There he is.’

Table 5.14: NIG Demonstratives and Presentatives

	m. sg.	f. sg.	m. pl	f. pl
Proximal	'da	'di	'do:l(a)	'de:l(a)
Distal	'dʕa:ka	'dʕi:ka	'dʕo:lak(a)	'dʕo:lak(a)
Presentative (prox)	'dawa	'diya	'do:lawā	'de:lāya
Presentative (dist)	'dʕakkuwa	dʕi:kkeya	dʕolakka'humma	dʕelakka'hinna

The origin of this *dVk* form is probably a demonstrative originally, but even outside of Egypt it is said to function as a referent presentative more than as a demonstrative. In the dialects of the south-central Hijaz, Theodore. Procházka Jr. (1988, p. 226) notes a form *duk* ‘look here’, and which can be optionally conjugated

⁸⁷ Note that these are representative samples of the forms and are not a comprehensive listing.

based on the gender of the interlocutor (and not on the gender of the indicated object) as *duk*, *dutʃ*, *dukum*, *dukin*, ‘f’or masculine, feminine, masculine plural and feminine plural respectively.

There is also something decidedly odd about all of these Nile Valley distal demonstrative forms, regardless of their origins. The sheer diversity of these forms is astounding (see Behnstedt and Woidich, 1985, maps 164–169), especially given the geographically one-dimensional nature of the Nile Valley. We find, for example, on map 169 in Upper Egypt thirty separate paradigms attested in an area representing only about 300km of river, with the two towns of il-Ḥimidāt (point 740) and il-ʿAšrāf (point 741) separated by some 8km having the following rather divergent paradigms (m., f., pl.): il-Ḥimidāt *dukkahi:ti*, *dikkahi:ti*, *dukkumʿmʿa:ti* vs. il-ʿAšrāf *dakha*, *dikha*, *dukhumʿmʿa*. The agglutinative nature of many of these demonstratives, for example the bizarre and multi-layered *dakhayyi:ti*, *dikhayyi:ti*, *dukhumʿmʿayyi:ti*, speaks to ongoing and recent change, especially given the relative transparency of these forms. Furthermore, these forms are extremely socially marked, and one of the first linguistic accommodations that immigrants from upper Egypt to Cairo make is to stop using these distal forms (C. Miller, 2005). This all suggests that these distal demonstratives forms are a strong site of social identity marking, and that rapid innovation of new forms may be a result of desire to mark group differentiation (or to cease marking group differentiation as in the case of migrants to Cairo).

Also, while these *dVk*-forms are listed as distal demonstratives in maps and atlases, they are vanishingly rare in actual usage (Egyptian Arabic is said to have a ‘one-term demonstrative system’ in Anderson and Keenan, 1985), and can be considered to be essentially separate from the demonstrative series. For example,

Mitchell (1962) notes that this series is primarily used to mark contrast as in (44), but compare (45) where an explicit contrast is made by only using near demonstratives. Davies (1981, pp. 171-176) notes that even in the seventeenth century, there was no true binary opposition between the near and far deictic series in Egyptian Arabic.

(44) (EGCAI, Mitchell, 1962, page no?)

muf da *lakin dukha*
 NEG DEM.NEAR.M.SG but DEM.FAR.M.SG

Not that one, THAT one

(45) (EGCAI, Hinds and Badawi, 1986, p. 273)

?inta ?aayiz da *walla da*
 you want DEM.NEAR.M.SG OR DEM.NEAR.M.SG

Do you want this one or that one?

Thus, the scenario in the Nile Valley dialects is that the demonstrative system appears to have largely lost the use of distal demonstratives. At the same time, these dialects have retained forms which might have been derived from the distal demonstrative originally, but which was used in presentative type constructions, which are often accompanied by pronoun suffixes. When a contrast is called for in these dialects, as in (44), these (essentially) presentative constructs are employed as distal demonstratives. Eventually they might have taken the place of distal demonstratives entirely, replacing the previous forms and creating these difficult to explain demonstrative forms.

5.5 Linguistic Chronology of Arabic Demonstratives

In this section I will attempt to explore the relative chronology of the developments which occurred in Arabic demonstratives. This section will rely almost entirely on linguistic data — the following chapter will incorporate social and historical data, but here we will try to keep the linguistic data separate to avoid circularity in the socio-historical reconstruction.

The earliest, and by far the most complete, split was between dialects with consonant-alternating demonstratives and vowel-alternating. This split was abrupt enough that we could even visualize it as a tree-like split. It is only in the branch which developed vowel-alternating demonstratives that we have all of the innovations related to the grammaticalization of **ha:* — the use of **ha:* initial demonstratives in consonant-alternating dialects is clearly the result of later borrowing and extension of the borrowed forms.

The innovation of diminutive demonstratives does seem to be limited to Arabic, but based on inherited Semitic material. It was probably developed before the split between consonant and vowel-alternating dialects, since it is found in both, though it seems least frequent in dialects with **ha:*-prefixed demonstratives.⁸⁸

The situation for the plural demonstratives is not clear — as discussed previously, it seems likely that two forms were inherited, /ʔula/ and /ʔila/. If we assume the simplest explanation for the current distribution of these dialects, that the association of these forms with particular genders is a much later innovation, then we can assume that very early Arabic dialects had the demonstrative forms shown

⁸⁸ It is not clear whether the Egyptian form *diyya* reflects the diminutive, or whether it is based on **ǰa-hiyya*.

Table 5.15: Early Arabic dialectal demonstratives

	Proximal	Distal
c. sg.	ǧa:	ǧa:k(a)
c. pl.	ʔula~ʔila	ʔulak(a)~ʔilak(a)
dim. c. sg.	ǧa:ya	ǧa:ya:k(a)
dim. c. pl.	ʔulaya~ʔilaya	ʔulaya~ʔilaya

in Table 5.15 Note that since dual demonstrative forms are not found in any modern dialect, we are not positing any dual forms. These were either restricted to the ancestor of Classical Arabic and lost by the time the ancestor of the dialects developed, or were a secondary innovation in that variety.

There is in principle no way to chronologically order the changes that created the dialects with *ǧa:/ǧi:/ǧu:* demonstratives relative to other innovations in the plural. This could have simply been a more extensive version of the same change that created the vowel-alternating dialects, which means that it was quite early, or it could potentially have happened at any time in the development of the dialects. Nor can the *ǧVk* presentatives-cum-demonstratives which are derived from those demonstratives be dated easily.

The same is true of dialects where the plurals are marked with the endings derived from the *-um/-in* pronominal markers — though both groups are “innovative” with respect to earlier states of the language, we cannot reliably date when those innovations occurred on purely linguistic grounds.

The most important development after the gender-marking split is the development of **ha:-*prefixed demonstratives. This development seems to have preceded the development of the adnominal **ha:-*prefix, since e.g. North African di-

alects have **ha:-*prefixed demonstratives, but not the adnominal **ha:-*prefix.

It is during the development of the **ha:-*prefixed demonstratives that we have the attachment of **ha:-* to the plural demonstratives, producing *hawla* dialects. Only after this can we get demonstratives like *ḍa:la*, *ḍayla*, and *ḍawla*. The continued diffusion of the **ha:-*prefix would result later in *haḍo:l* < **ha:-ḍa:-ʔula*.

The demonstrative *haḍanna*, since it is probably derived from *ḍalla* would therefore be later than the underlying *ḍalla* form.

The dialects which underwent the innovation that started with *ha:k* and ultimately resulted in the creation of *haydi* demonstratives have the complex internal chronology discussed above. They may potentially represent an earlier stratum of dialects than those with *haḍo:l*, since they have the linguistically earlier *hawla* demonstratives, but the two sets of dialects may well have developed independently in different areas, so we cannot give a language-internal chronology here.

The pronominal demonstratives derived directly from **ha:* including the elaborated form *ha:y* must have developed after the development of the *ha:-*adnominal determiners, since as far as the data shows, only dialects with *ha:-* have the pronominal demonstratives *ha:* and *ha:y*. The form *ha:y* must have developed prior to Stage 4 of the *haydi* demonstrative development process.

In a general schematic form, we can look at the developments in the following order, where the order is based primarily on language internal linguistic data, and where changes apply to some, not all dialects. See also the diagram in the following chapter, Figure 6.4 on page 418.

1. Leveling of *ḍa:* as a common masculine form

2. Split into consonant and vowel alternating dialects
 - From this point, all innovations occur in vowel alternating dialects
 - Dialects with plural *ǝu:* develop here at earliest. Presentatives with *dVʔ* would possibly develop starting here.
 - Dialects with plural forms derived from pronominal *-um/-in* affixed to *ǝ-* develop here at earliest.
3. **ha:* developed into a prefix which is affixed to **ǝa:* demonstratives
4. *ǝ-a/ay/awla* dialects are developed in analogy with *hawla* dialects
 - A further innovation seems to be the reinterpretation of *-awla* forms as masculine plural, and *-a(y)la* forms as feminine plural.
 - Continued expansion of **ha:-*prefixed demonstratives subsumes dialects which have innovated *ǝ-aw/ay/ala* demonstratives.
 - Dialects with *haǝanna* develop at earliest here.
5. **ha:* develops into an adnominal-only demonstrative in some dialects
6. **ha:* develops into a demonstrative pronoun
 - **ha:* is elaborated into *ha:k(a)* and *ha:y*
7. Some dialects with *ha:ka* undergo a complex series of changes, leading to the creation of *hayǝi:*

5.6 Conclusion

In this chapter, I have provided a linguistic history of the development of the demonstratives in Arabic dialects, based on a typological understanding of demonstratives, and situated within the history of the Semitic language family.

I argued that Arabic dialects experienced an early merger of masculine and feminine demonstrative forms, which forced dialects to re-invent marking to mark the demonstratives for masculine and feminine gender. Some dialects innovated demonstratives with vowels marking gender, while others innovated demonstratives which differed in their initial consonants. I explained further changes which transformed the plural demonstratives in these dialects.

I then show that among the vowel-alternating demonstrative groups, an inherited attention gathering particle **ha:* developed in various ways, first becoming a presentative before fusing with the inherited demonstrative forms to create a new set of demonstratives of the form *ha:ða:*, innovating into an adnominal demonstrative which primarily had a recognitional function, and finally into a pronominal demonstrative. These changes triggered a series of other changes in the demonstrative system in Arabic, some of which were quite elaborate.

Finally, I argued that the final *-ha:(t)* suffix that is widespread on demonstratives in Nile Valley dialects of Arabic is not the same **ha:* which acted as a prefix in Arabic dialects. I show instead that the apparent demonstrative forms it is attached to are actually presentatives, and in keeping with the general tendency for presentatives to have a following pronominal suffix, as well as based on the actual forms of the suffix, this form is actually a general pronoun referring to context and not cognate with **ha:* presentatives.

Chapter 6

Social, Linguistic, Historical Analysis and Reconstruction

This chapter draws together the history of the Arab migrations, the framework developed for understanding nomadic-sedentary interactions, and the historical linguistic analysis of Arabic demonstratives to understand the socio-linguistic groupings which existed pre-Islamically, as well as to explore how they developed following the Islamic conquests.

Since this dissertation only analyzes the demonstratives, the statements we are making here only apply to the speech communities in which these changes occurred in the demonstratives. It is likely that these speech community boundaries also correlate with alignment events that changed other parts of speech or linguistic forms, but we will not be able to discuss this in any depth here. Where there does seem to be a clear parallel to other important dialect features, we will make note of it in footnotes.

This chapter is organized by demonstrative form rather than by modern geographical distribution of dialects, though the two are related. It begins with consonant-alternating dialects (found primarily in Yemen today), and then dis-

cusses the *hawla*-dialects, which are wide-spread throughout the Greater Arabian Peninsula, with traces in Andalusian Arabic. The chapter then discusses the *ḏawla* and *ḏayla* dialects, which are represented today in Egypt and appear to represent a later layer of settlement in the Greater Arabian Peninsula than the *hawla*-dialects. The last major group of dialects are the *ha:ðu*-group whose origins are unclear, but which today are most widespread dialects in North Africa. Finally, the chapter discusses the unusual case of the Maltese and Classical Arabic demonstratives. By way of summary, Section 6.7 classifies all of the dialects found in the sample into the groupings described here, and Figure 6.4 is a visual representation of this information.

6.1 Consonant-Alternating Dialects

The fundamental split between consonant and vowel-alternating dialects appears to have been one of the very earliest events dividing the Arabic speech communities pre-Islamically. This split seems to have been deep and lasting — the consonant-alternating dialects participate in very few of the later changes which vowel-alternating dialects participated in, especially the attachment of *ha:-* to demonstratives. Consonant-alternating dialects which have acquired *ha:-*-prefixed demonstratives are generally marginal, and typically only develop the prefix after borrowing the demonstratives from vowel-alternating dialects.

Consonant-alternating dialects also typically appear to show **ʔila* as their original plural demonstrative, while the *hawla* dialects, and indeed most of the vowel-alternating dialects, show **ʔula*. This suggests that the vowel in the plural was an early shibboleth that marked two different speech communities — that is to say that the vowel difference in the plural differentiated two speech communities even before new feminine demonstratives were innovated to replace

common singular $\delta a:$. The two strategies for differentiating the common singular — consonant alternation and vowel alternation — each spread within those already differentiated speech communities. Thus the group of dialects using $*ʔila$ became consonant-alternating, while the group of dialects using $*ʔula$ became vowel-alternating, though dialects on the border between those communities have changed their demonstratives at different times in their history.

The range of consonant-alternating dialects is quite limited today. These dialects form a strip along the Yemeni Tihama (see Figure 5.1), but our data is not complete enough to understand their distribution further north in the central Saudi coast. Certainly within Jordan, and in all dialects for which we have attestations in the eastern half of the Arabian peninsula (including e.g. Hadramawt in south-eastern Yemen), we only have attestations for vowel-alternating dialects. Quranic Arabic and Classical Arabic grammatical texts attest a distal series with consonant-alternating singular demonstratives $\delta alika$, $tilka$ ¹, and as discussed in Section 5.3.1.1, the relative forms appear to be derived from consonant-alternating forms also.

However, consonant-alternating dialects had a wider distribution, particularly in the Levant. One example of this found is the Namāra inscription, written in 328 CE, and found some sixty miles southeast of Damascus (the most recent discussion of which is in Al-Jallad, 2012). It begins with the feminine singular demonstrative τy which could be read based on the modern dialects as either $ti:$ or the originally diminutive form $tiyya$.²

¹ Note however the somewhat unusual $*ʔula$ demonstrative $ʔula:ʔi$, which as mentioned above is more typical for vowel-alternating dialects. Furthermore, the distinction between the vowels in the masculine (/a/) and feminine (/i/) might reflect influence from vowel-alternating dialects (influence which is absent from any attested modern dialects).

² Rabin (1951, p. 152) suggests tai , which seems unlikely, since such a form is not attested in any

The evidence of the Namāra inscription suggests that at least one dominant dialect in the Levant was consonant-alternating. It is also possible that the dialect used in the Namara inscription is like that of Yafi^c (Vanhove, 2004), where the demonstratives have shifted largely to vowel-alternating demonstratives, but where a reflex of *ta:* is used in restricted contexts as a presentative and for other purposes, regardless the gender of the presented object (syntactically the predicate). However, the fact that this demonstrative agrees with the following noun NPS, cognate with Arabic feminine *naḥs* suggests that this was a normal use of a demonstrative in subject position agreeing with its predicate.

We also have some scattered attestations in Levantine dialects where *ta:* or similar forms are used as a demonstrative or relative (often derived from a demonstrative). In Cypriot Arabic, *ta:* is used as the primary relative marker, suggesting that at some point in its history, this dialect had consonant-alternating demonstrative forms which grammaticalized into relative markers, much as they did in whatever dialect Classical Arabic drew its relative markers from (Borg, 2004, p. 194; cited in Al-Jallad, 2012, p. 318).

Similarly, the Lebanese dialect of Baskinta seems to have apparently demonstrative or relative uses of *ta:* It is used in the phrase *kirma:l/minfa:n ta* (in.order.to that) ‘so that’, where it seems like it originally functioned as a pronominal demonstrative, though this has since become a frozen conjunction (Abu-Haidar, 1988, p. 80).³ In what may be a relative use, the *ta* particle is used with the interrogatives *layf* ‘why’ and *ki:f* ‘how’. The word *layfta* can only be used preceding a full

modern dialects in my sources.

³ Though Abu-Haidar (*ibid.*, p. 80) does not give any textual examples, I think this construction may be equivalent to *minfa:n he:k* in Damascene Arabic, where the second element is a modal demonstrative. It could however be a purposive clause of some sort, in which case this use of *ta* would ultimately derive from *hatta*.

clause, but *ki:fta* ‘why’ can be used as a single word question (Abu-Haidar, 1988, p. 116). This difference in usage, and the fact that *ki:fta* shows the meaning ‘why’ instead of ‘how’ suggests that the relative function of *ta* is secondary, with the original function as a pronominal (possibly discourse) demonstrative, which gives us a clearer reason to obtain the meaning ‘why’ < ‘how is that?’⁴

Shachmon (2013) explores demonstratives of the type demonstratives *huta* and *hita* in rural Palestinian dialects, and these could potentially have arisen from constructions like e.g. *hi ta* ‘she is that one’, equivalent to the similar *huwa ḏa* > *huḏa* used as a presentative in Sana‘i Arabic (*hu:-ḏa:-ħne:?* (he-this-us) ‘Here we are’ (W. Fischer, 1959, pp. 183-184). We find similar developments in Christian Middle Arabic texts, from (*ha:*) *huwa ḏa:* ‘there it is’ being used as a general sentential presentative, as in *ha: huwa ḏa: ana: ajlis* ‘Behold, I shall sit’, to losing the initial *ha:* and becoming a general presentative phrase *huwa ḏa: mraʔtuka* ‘there is your wife’ (Bloch, 1986, §6.3.1). Shachmon (2013, p. 76) does find that the *huta*-type pronouns are often more emphatic and are sometimes explained by native speakers as presentatives, but these forms could also simply be retentions of earlier Semitic pronoun forms with /t/ elements.

To summarize the evidence: Modern consonant-alternating dialects currently occupy at least the lowland, coastal plain of the southwest Arabian peninsula, though the rest of the Hijaz is currently terra incognita. The Namāra inscrip-

⁴ I accept that these are probably traces of an original *ta:* demonstrative, and I find the etymology Al-Jallad (2012, pp. 316-17) gives for *hatta* ‘until, in order to, so that’ < *ħa:d-ta:* (border-that) ‘to that point’ intriguing. However, I do not agree that the use of *ta:* as a purposive marker is directly derived from *ta:* — the use of that particle is basically identical to that of *hatta*, and deriving the purposive *ta:* from *hatta* via phonological erosion due to grammaticalization is a more parsimonious derivation than trying to propose a more complex series of events that lead a pronominal demonstrative to act as a purposive marker directly (see also Cowell, 1964, pp. 353,358). There are also clear traces in the Iraqi dialect of Khawetna of both the full *hatta* and reduced forms *hət* and *ta* with the same meaning and syntax (Talay, 1999, p. 190).

tion shows that a consonant-alternating dialect was the prestigious language of an Arabic speaking ruler who died in the Hawran in 328 CE, and Classical and Quranic Arabic both show traces of consonant-alternating dialects. Moreover some modern dialects in the Levant which are now of the vowel-alternating variety also show various traces of **ta:* used as a demonstrative (assuming that relatives are generally derived from demonstratives).⁵ These traces suggest a scenario of shift or diffusion, where the in-situ consonant-alternating dialects slowly shifted to become more similar to the *hawla* dialects (and other dialects) which moved into the Levant later on.

This analysis is similar to one possibility considered by Al-Jallad (2012, p. 323) who notes that “[*ta:*’s] absence in modern dialects could suggest that they have collectively lost it”; however he also suggests an alternative possibility, that “[*ta:* as a demonstrative] is an innovation that occurred in the dialect of Namāra, which is not directly ancestral to any of the modern dialects [in the Levant? - AM]”. Given the dialect data in this present study, it seems far more likely that the Namāra dialect providentially records for us a time when the distribution of consonant-alternating dialects were more widespread. The preservation of traces of consonant-alternating dialects also indicates that the dialect used in the Namāra inscription is almost certainly one (though not the only) ancestor of modern Levantine dialects.⁶

⁵ These dialects, particularly the Palestinian dialects, seem to exemplify the ‘edge of the steppe’ dialects which were theorized to exist in Section 2.5 — they are close enough to the steppe that they absorbed, at one point, the features present there, but then moved further in-land, so they were insulated from some changes emanating from the steppe and thus preserve some older features.

⁶ Al-Jallad (*ibid.*, p. 323, fn. 371) goes on to state that, if the scenario of independent innovation of *ta:* in the Namara inscription was correct, “this view would imply that several modern Yemeni dialects would descend from the Namāra dialect!” This statement underlines some of the problems with his model, in which modern Arabic dialects are formed by a process of linear descent

We cannot, at this time, resolve the question of when the current Yemeni consonant-alternating dialects came to be present in the Yemeni Tihāma, since we have a serious lack of data on dialects of west-central Saudi Arabia. The Tihāma should represent a spread zone, due to the relatively easy terrain, but it is also agriculturally rich, supporting a dense population. We hear that Arabs aided the Abyssinians in the Tihāma in the second and third centuries, so they may have settled there at that time (see Section 3.3.3). It is also possible that they moved south at a later date — perhaps the *hawla*-groups pushed the consonant-alternating dialects to move further south, eventually ending up in the Tihāma, though most of our reports of movement after the fourth century are northward not southward.

It may not be coincidental that the the fourth century CE is both our first attestation of an Arabic dialect, and the end of our attestations for Safaitic and ANA. It was perhaps the cultural ascendancy of the Arabic groups represented by MR? L-QYS that led to the end of the Safaitic inscriptional culture near the Levant. The Namāra inscription's symbolic use of a clearly Arabic language as a monumental language of inscription (though in Nabatean script), an example that is not reproduced again for almost two centuries, may represent a sort of cultural victory of Arabic over its ANA neighbors, though of course this is all necessarily quite speculative.⁷ Nonetheless, the historical accounts of movement into the Levant

from Levantine Arabic dialects. The dialect geography, with consonant-alternating dialects well established and very diverse in the south-western coast of the Arabian peninsula, would instead point instead towards a (more) southern homeland for consonant-alternating dialects. Moreover, if one really is looking for a historically deep, proto-Levantine Arabic, to ignore the dialect recorded in the Namāra inscription would be ignoring an important stage of the development of Levantine Arabic, and almost ensures that with respect to the demonstratives at least, a reconstruction would actually capture a stage of language different from the proto-language of the (equally Levantine) Namāra inscription.

⁷ To go deeper into speculation, one wonders whether MR? L-QYS felt a need to write in Arabic specifically to differentiate himself from the remnants of the Safaitic culture, a need which would not be felt by later dynasties.

around this time, and the apparent contemporaneous end of Safaitic culture suggests strongly that movement around the third or fourth centuries brought Arabic (in the form of consonant-alternating dialects) into the Levant. We cannot, however, determine whether consonant-alternating dialects hailed from further south, or had split off from their southern relations for some time, or whether these dialects were primarily found in the north in the fourth century, with later southern groups branching off of a northern group of dialects.⁸

6.2 *Hawla* dialects

The first emergence of **ha:-*prefixed demonstratives in Arabic dialects appears to have occurred within the group of vowel-alternating dialects, and would have resulted in dialects whose demonstratives looked like *ha:ða:*, *ha:ði:*, **ha:ʔula* > **ha:wla* (> *hawla*), and for that reason I will call this the *hawla* group, to distinguish it from later groups which developed plural demonstratives such as *haðo:la*.

Historically, it is not entirely certain when this group arose. In ANA, the Semitic language closest to Arabic, **ha:-*prefixed demonstratives were found quite early. In Dadanitic, written in the oasis of Dedān in the northern Hijaz until the 1st century CE, we find the phrase H-ṢLMN ḤDH (def-statuettes this) ‘this statuette’, in the inscription JSLih 82/1, cited in Macdonald (2008a, p. 209).⁹ Fares (2003) gives a probable date for this inscription in the last third of the fourth century CE (i.e. 330s BCE). The **ha:-*prefixed demonstrative here, however, may resemble the development in Hebrew where there is agreement between the article and

⁸ For further discussion of the ‘origins’ of the various Arabic dialects, specifically the ‘out of Levant’ or ‘out of Yemen’ hypotheses, see Section 6.9 in the Conclusion.

⁹ The demonstrative was probably *ha(:)ða:*, though we do not know what the feminine equivalent was.

the demonstrative rather than the development of **ha:-*prefixed demonstratives in Arabic where **ha:* became attached to the demonstrative in its function as a presentative. Nonetheless, we can see that there were **ha:-*prefixed demonstratives present in the linguistic environment of the Arabian peninsula as early as the fourth century BCE.¹⁰

The *hawla* group of dialects seems to have been a sociolinguistically very cohesive group. We have argued in Section 5.3.2.1 that dialects with **ḏa-ʔila-*type demonstratives actually calqued those demonstratives from the *hawla* dialects. What is surprising is that that calqueing process cut across consonant-alternating and vowel-alternating dialects, and happened in different ways within those dialects. This suggests that the contrast at the time wasn't between consonant and vowel-alternating dialects per se, but between *hawla* dialects and everyone else. The nature of this alignment event is also interesting, since it represents a kind of contact-induced remodeling rather than a diffusion event, further suggesting that *hawla* dialects were a cohesive and distinct social group whose dialect was seen as foreign by those adjusting their own dialects.

We have enough evidence to suggest a possible homeland for the *hawla* dialects. The Dadanitic example comes from the northern Hijaz, and that might suggest a fairly northern origin for the spread of **ha:-*prefixed demonstratives (i.e. from ANA to Arabic). However, the *hawla* group is extremely deeply embedded in the highlands of Yemen, as shown in Figure 6.1. This difficult, rugged terrain would have limited contact between speakers, and thus would have taken a significant amount of time to shift from the Old South Arabian languages to Arabic. This suggests that the *hawla* group has been very long established in the Yemeni highlands.

¹⁰ For a sense of scale, Dedān is some 750km SSW of Namāra, and some 600km north of Mecca.

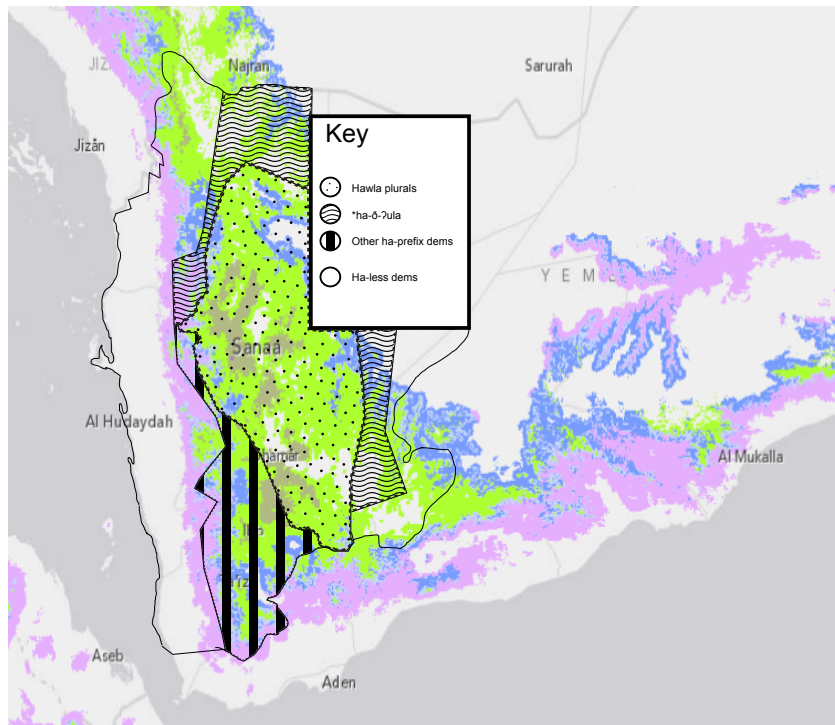


Figure 6.1: Plural demonstrative forms in Yemen

However, since contact was necessary between *hawla* dialects and non-*hawla* dialects to produce *ḏa-ʔila* type demonstratives (see Section 5.3.2.1), there must have been a phase when the *hawla* and non-*hawla* dialects were in geographical proximity. At first glance, this would certainly appear to be the case in Yemen, but we also have the testimony of Al-Hamadani that OSA languages were still spoken in the highest mountains, between the *hawla* and non-*hawla* groups by the tenth century (see Figure 3.3 on page 165). If this portrayal is indeed accurate, it means that there would have been non-*hawla* dialects on the coasts and *hawla* dialects in the highlands, sandwiching the surviving OSA-speaking areas between them, and hence not necessarily in contact.

This would suggest a homeland for *hawla* (and for the development of *ḏa-ʔila* dialects) further to the north where there would be no such interposed languages, which could be as close as the areas immediately north of Yemen, near e.g. Najran. This again is an area where we need significantly more data before we can make any decisive statements.

Pin-pointing the time when *hawla*-dialects first developed is difficult, and extremely speculative. We have multiple phases of Arab migration into Yemen, often apparently from further north in the Hijaz into the Jawf region, with the most important and deeper penetrations happening after the turn of the era, and by the third century we have reports of mountain dwelling Arabs (see Section 3.3.3). The historiography of Omani sedentary dialects suggests that these dialects are descended from tribes which emigrated after a collapse of the Marib dam in the second century, though this may not be entirely accurate (see Section 3.3.4). Older Omani dialects are not of the *hawla* type, but rather from the *ḏayla*-type dialects. Since their ancestral dialects must have calqued the *hawla* demonstratives, it seems likely the *hawla* dialects were already present in the area, if not in the Jawf itself before the Omani diaspora. If we accept the Omani history, then the alignment event which created *hawla* dialects occurred before the second century CE not far from northern Yemen, but far enough from the Jawf to be unaffected by the second century collapse of the Marib dam.¹¹

In the modern era, dialects of the *hawla* type are well established west of the major cities of Syria, including dialects in what is now Turkey to the NW of Syria, so-called Cilician Arabic (S. Procházka, *EALL* : “*Cilician Arabic*”). They are

¹¹ This is significantly later than the 4th century BCE date given to the similar looking Dadanitic $\text{H}\mathbf{\bar{D}}\text{H}$ form, but that form and the Arabic form are probably the result of very different pathways of development, and so they are not necessarily related.

also well established in dialects located between the Tigris and Euphrates, whether in Anantolia or Iraq, what are typically referred to as ‘qəltu’ dialects (Jastrow, 1978). The *hawla* dialects are also well established in modern Yemen as shown in Figure 6.1. Almost all of the *hawla* dialects outside of Yemen show developments related to the innovation of *ha:k(a)* distal demonstratives, as described in Section 5.4.4.1, suggesting that early ancestors of these dialects had already innovated at least the *ha:k(a)* distals by the time they split off from the southern *hawla* group.

6.2.1 *hawla* dialects in the Levant

When did these dialects come to the Levant? As we’ve seen, the Namāra inscription records at least one dialect with consonant-alternating demonstratives in the Levant in 328 CE, so they probably arrived in the Levant after that time. There is also an inscription at Ḥarrān, near Damascus, which shows a masculine singular demonstrative *ḏ?*, and which was written in 568 CE (Al-Jallad, 2012, p. 121). This could be an indication that **ha:-*prefixed dialects had not yet become thorough entrenched, or it could also be a dialect in which **ha:-* was still optional as a prefix. It could even reflect the consonant-alternating dialects. It does not, however, tell us what dialects further south were like.

As discussed in Section 3.3, epigraphic data as well as the post-Islamic historiographic record all point to a series of northward movements out of the Arabian peninsula and into the Levant and other regions. The first tribe that we hear of as a client state of the Romans is that of Ṣalīḥ, who is said to have moved from the Saudi Tihama but they are said to have moved into the area not from the Hijaz, but from Wādī Sirḥān, a valley that cuts from near Damascus through the Hawran to the south-east, into the peninsula (See Section 3.4).¹²

¹² There are also the claims by Shahīd (1984) that the Namāra inscription represents a Tanūḥīd

Another group that may have brought *hawla* dialects into the Levant are the Ghassanids. They are said to have been under Himyarite rule originally, which would accord well with the original homeland of *hawla* dialects just north of Yemen. Independent epigraphic data records the Ghassanids as living in the central Hijaz in the third century, and they were known to be clients of Rome by the sixth century, having ousted Ṣalīḥ. If the Ghassanids were the bearers of *hawla* dialects, that would give a very early sixth century date for the entrance of *hawla* dialect into the Levant. Such a date would accord reasonably well with the attestation of the non-*hawla* demonstrative form in the 568 Ḥarrān inscription, since it might have taken some time for *hawla* dialects to become dominant, especially as a language of prestige and writing.

An alternative theory is that of Al-Jallad (2012, pp. 25-28), who argues that the original homeland of Arabic was in the Levant, and that this must be the *Urheimat* from which other Arabic dialects later descended.¹³ It is possible that on a time-scale of millenia this is true, but the Arabic-speaking presence in the Levant, prior to the Islamic conquests, was likely restricted to the steppe-lands (see Section 3.4), which would have been easily repopulated, and so we do not necessarily have access to dialects which date to the time (see also the discussion in Section 6.9).

dynasty, though they presumably represented consonant-alternating dialects. However, exactly what his evidence rests on other than some outdated readings of the Namāra inscription itself is not clear, and it is difficult to disentangle his sources of evidence from his conclusions.

¹³ Among his arguments is that Levantine Arabic varieties show a greater diversity than Yemeni varieties. Measuring diversity is difficult at the best of times, but for demonstratives this is patently not the case. For the demonstratives, the amount of diversity in the Yemeni dialects far exceeds that of the Levant. Though both North Yemen and modern Syria are about the same size, almost all of the different demonstrative types reconstructed by this study are found in North Yemen, while the demonstratives found in Syria are almost all variations on *hawla* dialects and *ḍawla*, *ḍanna* dialects.

Al-Jallad (2012, p. 25) notes of course that in modern historical work, the notion of population replacement by “endless waves of nomads” has largely fallen out of favor, a point also made by Hoyland (2009). However, in areas as sparsely populated as the steppe, it does not take ‘waves of nomads’ to replace the population. Indeed a ‘repopulation’ of this area would not necessarily be noticeable to an outside observer, who might not necessarily be able to tell the difference between population replacement due to seasonal migrations from population replacement due to a forceful invasion of territory.¹⁴ The principle of first effective settlement, therefore, is not nearly as meaningful in lightly populated areas than it is in settled agrarian lands.¹⁵

There was of course more intensive settlement in the Hawran region, an area that contains both steppe and more fertile areas. However, this is the very area where the Roman sponsored client states tended to colonize, so here we really might expect a change in population as one military group ousts another, as reflected in the historical and historiographical records. Still, not everyone would leave -- much of the non-military population would certainly remain, and the continuity in this population accounts for the traces in the modern Levantine dialects of consonant-alternating dialects.

The Ghassanids would have moved into space occupied by the Ṣalīḥ, and they probably brought with them a large number of their tribesmen (whether immediately as part of a conquering army, or after establishing their dominance,

¹⁴ Imagine for example a dialectologist doing research in an area where different groups are present in different seasons. If they worked in the winter they might get a very different result dialectologically than if they worked in the summer.

¹⁵ Ease of population replacement in lightly populated areas is also behind the Bedouin Paradox, the dialectological problem that Bedouin dialects (in the sense of the dialects of speakers living in sparsely populated marginal lands) are often quite new to an area, but appear very different from nearby sedentary dialects (see Section 2.5).

via later migration). We know that the Ṣalīḥ remained in the Levant some capacity even after the Ghassanids came, so perhaps they were primarily ousted from power and perhaps the more desirable lands (Shahîd, *EI2*, “*Salīḥ*”), so we certainly have evidence of continuity. We do not know what dialect type the Ṣalīḥids spoke. The Ghassanids, speaking a *hawla* dialect, would have represented the language of prestige in the area, and so we would expect a slow shift to *hawla* dialects from the original inhabitants of the Levant, which is consistent with the preservation of traces of *ta:* in those dialects and even perhaps with the use of a **ha:-*less demonstrative in the Ḥarrān inscription.

The Ghassanids primarily colonized the Hawran and possibly transitional areas between steppe and agriculturally rich Aramaic speaking areas further south in Jordan. However, the *hawla* dialects are attested well inland in the Levant, and indeed they are no longer present in the Syrian part of the Hawran at least.

Key to understanding the dialect geography of these inland Levantine *hawla* dialects is understanding how the Levant was Arabized. As discussed in Section 3.5.4, the primary Arabizing force in the Levant was probably not the conquests of the cities per-se, but rather the devastating plagues that wracked the area from 542–749 CE. These plagues would have created a vacuum into which the relatively unaffected Arabic speaking nomads and semi-nomads could have moved, slowly replacing the former Aramaic speaking agricultural population. However, in the cities, even though the plague would have struck higher density areas more heavily, there would probably still have been substantial populations remaining.¹⁶ These urban populations which were likely still heavily Aramaic speaking, and so the Arabization of the cities would have taken much longer than the Arabization

¹⁶ That is to say, even a 25% mortality rate in a city leaves significantly greater absolute and relative population than in the countryside.

of the countryside, as a much larger and denser group of Aramaic speakers would need to shift. This scenario also suggests that the initial settling population of Levantine cities might not have been the ultimate Arabizing population, which may have come much later.

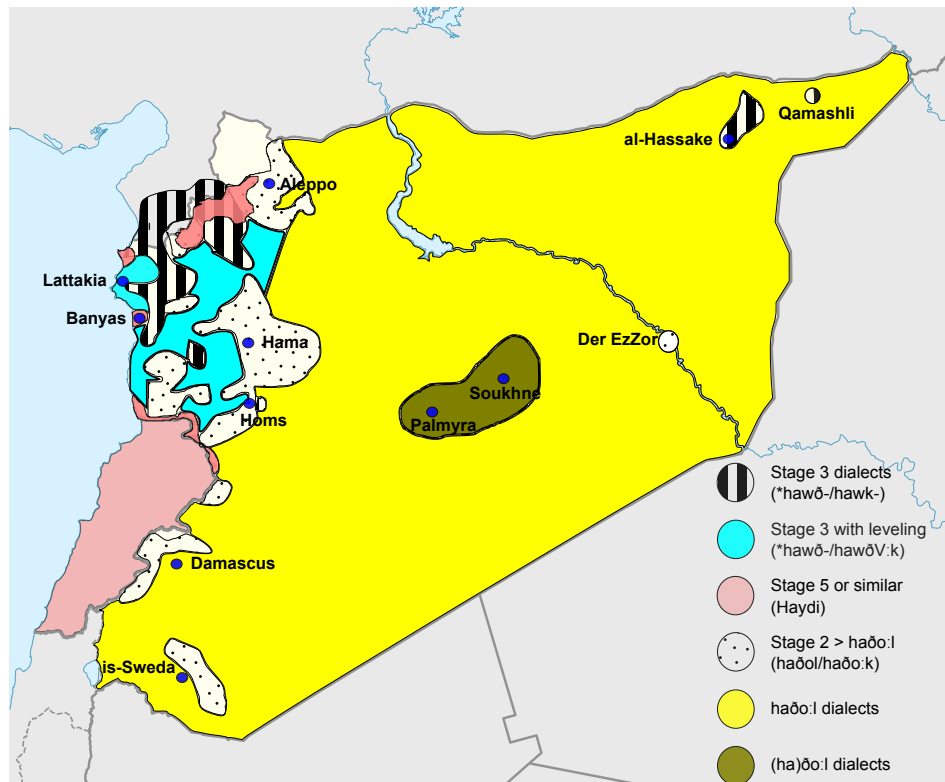


Figure 6.2: Dialects Syria by Stage of Development as described in Section 5.4.4.1

Given this scenario, and the dialect geography of Syrian as portrayed in Figure 6.2, the then-nomadic and semi-sedentary dialects of Syria must have been of the *hawla*-type, as are their descendants in the areas between the cities and the

coasts.¹⁷

This process potentially could have begun earlier, but was likely most effective following the beginning of the plague era in 542 CE. If indeed it was the population movements represented by the Ghassanids which brought *hawla* dialects into the Levant, then the earliest they could have been in the area was in about 490 CE, so they would have had about two generations before the plagues started in which to establish themselves in the area, which seems a reasonable length of time for them to establish their dialect.

There dialects in Cyprus are of the *hawla* type, as expected since they moved from the Levant to the island between the ninth and thirteenth centuries. These dialects do not resemble any of those described in Section 5.4.4.1, with proximal demonstratives *aḏa*, *aḏi*, *alli* (< **hawla*) and distal *aḏak~ak*, *aḏik~ayk*, *al-lik*. These unusual forms are probably due to a process of koineization which has happened over a period of centuries as the Maronite, Arabic speaking population of Cyprus dwindled. Karyolemou (In press) notes that fourteenth century sources report some sixty Maronite villages as of 1224, but by 1661 only ten villages are reported. The contraction of the population was accompanied by geographic contraction — where before, villages had been scattered across the island, the five villages that remained in 1878 when the British arrived were all concentrated in a contiguous area in the north. As speakers moved from many villages into only a few, a classical case of koineization would arise where different groups would

¹⁷ Indeed, Al-Jallad (2012, p. 325) reconstructs **ha:wle*: as the proximal plural demonstrative for his “Proto-Levantine Arabic”, which does reflect the most effective colonizing dialect in the area, with a note that perhaps **hawḏe* reflects a later development, which also is supported by the analysis here. Nonetheless, the time-depth of this “Proto-Levantine Arabic” must be quite shallow, and misses the variety of dialects that must have occupied the Syria deserts over time, including the consonant-alternating dialect of the Namāra inscription and the **ha*:-less dialect of the Ḥarrān inscription

be using different forms (which in this case, would represent the various stages of *hawla* dialect developments). Koineization tends to level to a single set of forms over time, with some being retained and some lost. The optional forms in the distal forms appear to vary between a very traditional *hawla* dialect, possibly one without *ha:k* distal demonstratives, and a dialect in which the *ha:k* distals were well developed, matching Stage 4 in the account here (shown by the feminine distal **hayk*). We expect to find a *hawð*-type plural demonstrative in Stage 4 dialects, but the dominant dialect(s) in the koineization process probably had **hawla* as the proximate plural demonstrative.

We also find *hawla* dialects in Andalusia, though this is expected given the large Syrian presence in Andalusia. The Syrian army dispatched to crush the Berber rebellions in North Africa, assembled in 74/694 and arriving in Andalusia in 124/741 (after a number of stop-overs) could certainly have brought speakers of a *hawla* dialect into Andalusia, though other dialects would probably have arrived with earlier migrations. This timing of this matches well our presumed penetration of *hawla* dialects in the Levant by the beginning of the eighth century due to the plagues. On the other hand, we have only minimal attestations of Andalusian Arabic which seem inadequate to show what must have been greater dialect variation (See primarily Corriente, 1977; Zaragoza, 2012). Nor can we rule out the possibility that Andalusian attestations of *HWLA* are not meant to represent Classical Arabic *haʔula:ʔ*, especially since most of our Andalusian data comes from elevated literary and artistic texts.

6.2.2 *hawla* dialects in Mesopotamia

The attestation of *hawla* dialects in north-eastern Syria's Mesopotamian regions, as well as in most of the *qaltu*-type dialects suggests that these dialects

occupied the essentially continuous desert region between the Euphrates and the agriculturally rich lands of the Levant. Though this could be the result of a single movement event, the historiographical tradition suggests that two different migrations could have brought *hawla* dialects to the northern part of the Peninsula. The Arabic historical tradition holds that the original founders of al-Ḥira in Iraq were also from Yemen, the Quḍā'a branch of the Tanūḥ federation, that they moved into the region in the third or fourth century, and that in their migration into Iraq, they passed through the eastern and north-eastern Peninsula, including the areas around Bahrain (see Section 3.6.1). This route sounds quite similar to the pre-Islamic trade route that ran from Najran to the north-east, through the Yamama, and then to Arabian Gulf coast at al-Qatīf, just across the Gulf of Bahrain from modern Qatar (see Kennedy, 2001)

A two-migration account would also explain the presence of apparent traces of the *hawla* type dialects in Bahrain, where Al-Tajir (1982, p. 101) notes that women in the village of Dirāz still use distal *ha:k(a)*, *he:k(a)*. This village traces its history back to the eastern Peninsula (Holes, 1987, p. 39), which would be consistent with a scenario in which some members of this tribe settled in eastern Saudi Arabia en route to Iraq, and then moved into Bahrain, where their dialect has preserved these early features, in spite of being heavily influenced by surrounding dialects. This would also explain how we find *hawla* dialects on both sides of the dividing line between two major world empires: these two migration events would have both occurred after the alignment events which created *hawla* dialects.

6.2.3 *ha:k(a)* developments

As discussed in Section 5.4.4.1, a subset of the *hawla* dialects also developed more elaborated **ha:-*prefixed demonstratives through a series of changes,

beginning with the first step of innovating *ha:k(a)* as a distal demonstrative for all genders a numbers. This section discusses whether these developments occurred in an earlier, ancestral dialect which gave rise to the dialects in Mesopotamia and the Levant, or whether these changes happened in-situ.

The *hawla* dialects in Yemen did not make the first step that leads to these changes, the innovation of far *ha:k* demonstratives. Though there are a small number of dialects which have forms of the type *hawða* in the area near the main *hawla* dialects in Yemen, these forms probably arose through a different process than the one we find in *ha:k*-dialects. For example, Yemen #120, immediately adjacent to the *hawla* dialects, has the forms shown in Table 6.1. Here, some sort of analogical process appears to have led to the remodeling of the proximal plural based on the distal plural, a process quite different from that found in the *ha:k* dialects.

Table 6.1: Demonstratives in Yemen #120

	Proximal	Distal
M.SG	haða	haðak
F.SG	haði	haðiʃ
C.PL	hawða	hawl ^ʕ a:k

In contrast to this, in the Levant, Mesopotamia and Anatolia, there are almost no dialects which shows the *hawla* ancestral forms (see Table 5.12 on page 344) — only the Sason dialects are at Stage 1, while the remaining dialects are all at stage 2 or more advanced. In the Levant, a small number of dialects at the extreme north and south of Lebanon are still from the ancestral *hawla* variety, but the vast majority of dialects are more advanced, with stage 3 and higher dialects the most widespread. We also have the evidence of Bahrain, where we find traces of *ha:k*

(i.e. Stage 1, possibly stage 4 forms since we find feminine *he:k* also) forms which are themselves probably derived from dialects on the Gulf coast.

This suggests the following scenario: The *hawla* dialect speech community arose somewhere in the south-western Peninsula, north of the core OSA speaking areas. One branch moved south, into Yemen, but another branch probably remained further north. It was that group that innovated the *ha:k* demonstratives, and probably splitting into two groups once it had reached at least Stage 1 in the development of the demonstratives. This scenario is supported by the historical sources, which state that both the Ghassanids and the Tanūḥ (later Lakhmids) who moved north were both at some point under the control of the Himyarites, i.e. in the south-west Arabian Peninsula. These groups then moved north, with the Tanūḥ supposedly in Iraq, founding al-Ḥira by the third or fourth century.

It is less clear whether the dialects that entered either Syria or Iraq were at Stage 1, at a later stage as a whole, or had a great deal of internal diversity, with some groups progressed to Stage 2 or further. The modern distribution of the dialects according to these stages can be seen in Figure 6.2 for Syria and in Figure 6.3 for Mesopotamia. What makes it difficult to determine how and where these dialects developed is the high probability of parallel by independent development: Each of the steps from Stage 1 to Stage 5 are logical analogical processes and are likely to happen. On the other hand, they are not inevitable. The dialect of Sason, for example, developed gender and number marking for the *ha:k* demonstratives by adding endings to them, so it has m., f., pl. *a:ga* < **ha:ka*, *a:gi*, *a:gu*, a very logical alternative to the Stage-wise development, which only produces full gender differentiation of the distal forms by Stage 5.

The complexity of these developments means that it is more likely that each of these stages was reached only once, and then distributed migration, rather

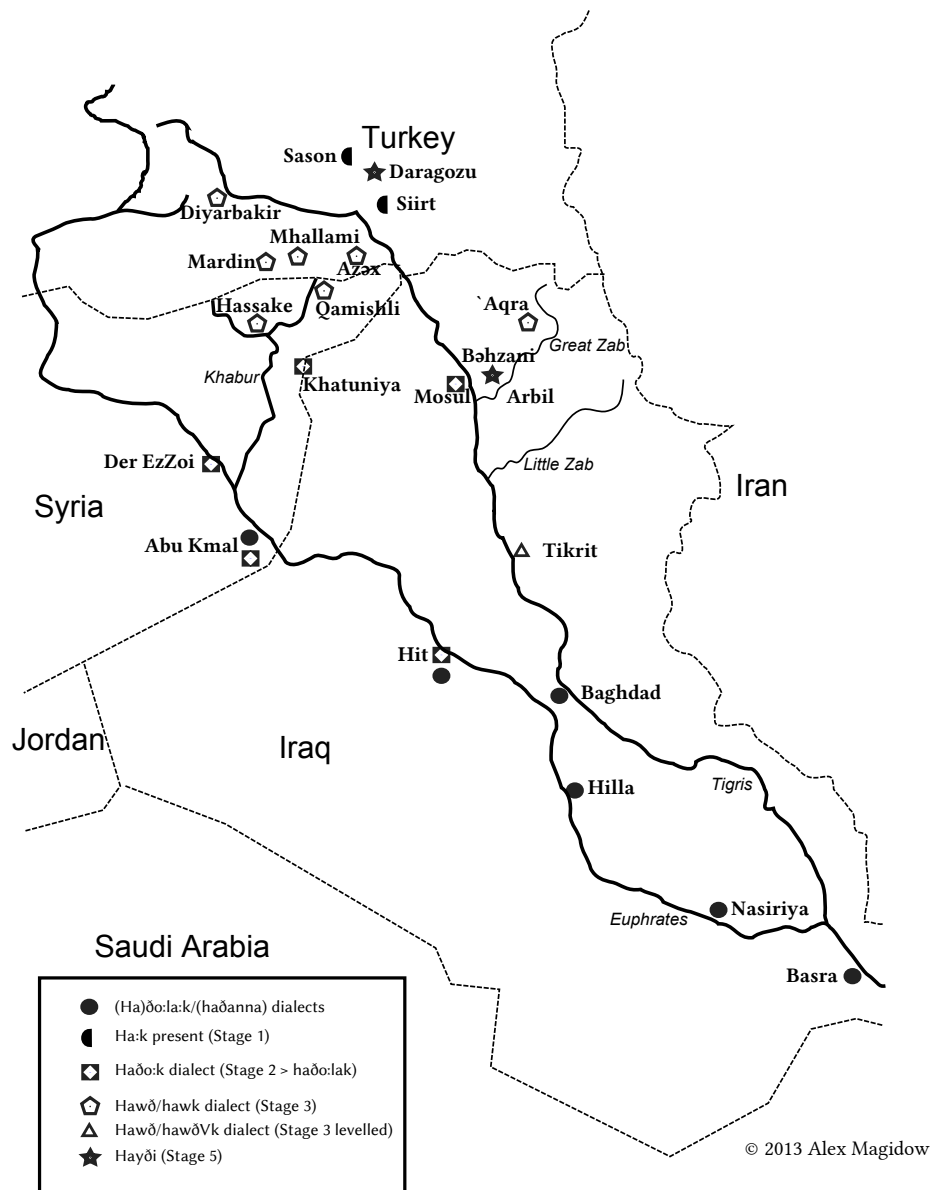


Figure 6.3: Map of Demonstratives in Mesopotamia.

than as a result of parallel evolution in-situ. This theory is supported by the apparently chaotic distribution of these dialects. The distribution of different Stage dialects do not show many clear areal patterns what we would expect from in-situ evolution. Consider how scattered Stage 5 dialects are, sandwiched between other dialect types in Bəhzani and Daragozu, in small groups in northern Syria, but well established in Lebanon.¹⁸

The cluster of Stage 3 dialects in northern Mesopotamia, from Hassake to Diyarbakir could perhaps be the result of areal diffusion. However, we do not find any Stage 3 levelled dialects (where distal plural *hawk* is replaced by extending proximal *hawđi* > *hawđi:k*), though we would expect some dialects of that type to be innovated spontaneously, especially as a result of in-situ diffusion. Instead, the distribution looks more homogeneous, suggesting a movement of peoples all using Stage 3 dialects from a center, possibly from Hassake up-river to Qamishli and then into Anatolia, or the opposite.

Stage 2 developments must have been widespread when *ha:đo:l* dialects came into the area, since the *ha:đo:k* dialects that result from contact between those two dialect types are widespread (see Section 5.4.4.1 for details on how that contact form arose), assuming that the *ha:đo:k* dialects are the result of parallel, independent innovation caused by very similar contact situations.

However, the *ha:đo:k* dialects might also have diffused from a single point or area of innovation — two paths seem likely, either from Syria (possibly Aleppo?) down the Euphrates, or the opposite direction, up the Euphrates from pre-Islamic cities which had Stage 2 dialects. Certainly the Euphrates likely used *ha:đo:k* more

¹⁸ Though Behnstedt's (1997) data on Lebanon is rather limited, and the decision to 'color in' Lebanon based on a small number of sample points is not necessarily good practice. The map here simply reflects the decisions made in that source.

widely in the past — the relatively small city of al-Bukamal still preserves traces of the *ha:ðo:k* dialect, though it is clearly shifting to the now more widespread *ha:ðo:l* dialects now found in the area, and the same is true of Hīt, which used to be a major stop along the trade routes between Baghdad and Aleppo. Both of those cities were always small, and probably were fairly low in the urban hierarchy, suggesting that *ha:ðo:k*-dialects were not necessarily ‘urban dialects’ in the Euphrates valley, though rural/Bedouin *ha:ðo:l*-dialects have since wiped away evidence of the earlier rural *ha:ðo:k*.

The direction of diffusion may have come from Iraq. Linguistic forms tend to move (counter-intuitively) from downstream to upstream in Iraq. For example, the Anatolian dialects of Arabic seem to be least influenced by changes from further downstream such as the adoption of *gəlat*-dialects. This would weight towards an ultimately Iraqi origin of the *ha:ðo:k* dialects but much more evidence is needed, however, before we can be anywhere close to certain.

The *ha:ðo:k* dialects seem to represent a type of calqueing, similar to the development of *ðawla*-dialects, discussed below. Instead of replacing their native plural proximal *hawl* and distal *hawk* with **ha:ðo:l*, **ha:ðo:la:k*, these dialects created an interdialect form based on proximal (donor-dialect) **ha:ðo:l* and a (recipient-dialect) understand of *-l* and *-k* as representing a proximal-distal binary, hence producing distal *ha:ðo:k*. This suggests that the Stage 2 dialects that came into contact with **ha:ðo:l* dialects were sufficiently prestigious and well-established (perhaps even numerically superior) that they resisted complete convergence to **ha:ðo:l*. This implies that at the time of contact between these dialects, Stage 2 dialects were relatively prestigious and important, speakers of those dialects viewed **ha:ðo:l* dialects as foreign, but speakers of those dialects had enough prestige or numerical weight to impose their demonstrative forms on the Stage 2 dialects. We

will discuss the sociohistorical implications of this further in Section 6.3.3.

The Stage 5 dialects (using **hayǿi* for the proximal feminine singular), as mentioned earlier, are quite scattered, apparently at random. The historical data on Lebanon, though, discussed in Section 3.5.5 finds that one colonizing group (the Tanuḥ, who became Druze) in Lebanon was drawn from northern Syria. If that group was from the Stage 5-speaking area, this could explain both the (somewhat marginal) attestations of Stage 5 dialects in northern Syria today, and the apparently wider distribution of Stage 5 dialects in Lebanon. Since this group gave birth to the Druze, we can test this hypothesis once we have more dialect data from Lebanese dialects.

6.3 Dialects with *ǿawla*, *ǿayla* plurals

Section 5.3.2.1 argued that the demonstrative forms with the *ǿa:* prefixed to the original plural demonstratives *?ula* and *?ila* were originally formed through contact with *hawla* dialects. This seems to have occurred in groups with both original *?ula* and *?ila* plural demonstratives, so we find both **ǿa:-?ula*, normally realized as *ǿawla* or with /aw/ > /o:/ as *ǿo:la*, as well as *ǿa:-?ila* > *ǿayla-ǿe:la*. Since these forms were created at different ways in different places, we also find *ǿa:l(l)a* forms, where the vowel of the initial *ǿa:* did not end up forming a diphthong with the initial vowel of the plural demonstrative.

It is not clear whether the dialects which underwent these innovations formed a speech community. The fact that this change is a remodeling of forms based on a ‘foreign’ dialect (instead of diffusion) does suggest that the *hawla*-using speech community was viewed as an out-group (see Section 1.3.3.1) by speakers of the dialects which developed *ǿVla* forms. This might suggest a binary opposition between *hawla* and non-*hawla* dialects.

On the other hand, a variety of forms were produced by this change, without a later diffusion of a single form throughout the speech community that would have hidden this variation from us. That diversity would point instead to a one (*hawla*-dialects) to many speech community relationship, where the $\delta a:-\text{?}u/ila$ dialects actually represented a number of different speech communities. We also have few examples where all of the δVla dialects act as a group later on, but many of the subdialects of this group do seem to operate in a more cohesive fashion. All of this suggests that the remodeling of the $\delta a:-\text{?}i:-ta:-\text{?}ila-\text{?}ula$ demonstrative series to more closely match that of the *hawla* dialects was actually a series of independent alignment events diagnostic of the existence of many different speech communities.

Indeed, this type of one-to-many relationship between a single well defined speech community and many smaller, but not cohesive speech communities, is suggestive of a scenario where one group dominates many others, a type of imperial-relationship. This would also explain why there was remodeling rather than diffusion — the dominated groups would be linguistically asserting their independence from the dominant group, while still under its political and linguistic sway.

Since dialects with $\delta ayla$ forms do not appear to have moved significantly beyond the Arabian Peninsula (represented here by AROM, ARQAUZ, both vowel-alternating, and Y96 a consonant-alternating dialect), particularly the southern half of the Peninsula, and since we know the *hawla* dialects also arose in the southern Peninsula, this suggests a southern Peninsular origin of the $\delta ayla$ and $\delta awla$ dialects. Again, if we take at face value claims that Omani Arabic first moved out of the Yemeni Jawf by the second century, this suggests that we should further investigate the history of dominant Arabic-speaking groups in the southern

Arabian peninsula around the second CE to better understand the circumstances surrounding the birth of these dialects. The fact that this change subsumed both vowel and consonant-alternating dialects again suggests that these forms post-date the split between consonant- and vowel-alternating dialects, which we can thus date to prior to the second century CE.

6.3.1 Free variation of *ḍawla* and *ḍayla*

Some dialects also appear to have been formed from ancestral dialects which belonged both to the *ḍayla* and the *ḍawla* groups, and this creates mixed dialects where variation between the two demonstrative types apparently is not meaningful. However, all three of the sources (for dialects ARBAH, ARANZ, YHADR) which report this variation cover dialects which subsume numerous villages and groups, so it might actually be an artifact of the research method (listing all of the forms found among multiple villages) rather than reflecting actual behavior (speakers using *ḍawla* and *ḍayla* truly interchangeably.)

Indeed, even if these sources are reporting accurate language use, the development of free variation in the use of these demonstratives could have happened independently in a number of places, and therefore this is not diagnostic of an ancestral speech community.

6.3.2 Gendered *ḍawla*, *ḍayla*

A much more important pre-Islamic speech community is the community which, for reasons which are not entirely clear, reallocated the *ḍawla*~*ḍayla* dialectal variation into gendered plural demonstratives, marking masculine and feminine plural respective (i.e. *ḍawla*, *ḍayla*). We can refer to these as gender-reallocated dialects, for which there are two important modern branches: One branch, which

had **ḏayla* as the feminine plural, moved into Egypt, where its descendants are widespread throughout the Nile Valley. The other branch, which had *ḏalla* (which became *ḏanna*) as its feminine plural moved into urban centers in Iraq and Syria.

This speech community, in order to perform this reallocation, was likely situated between dialects which had generalized either the *ḏawla* or the *ḏayla*. They were probably already of the *ḏawla* type, since it seems more likely that re-allocation of these forms would assign the foreign form to the “open” slot in their inventory, rather than switching out an already productive *ḏayla*, moving it to the more marked feminine use and replacing it with a foreign form *ḏawla*. If there was a more koineization-type situation, where the two dialects were suddenly brought together, there still seems to have either been a majority of *ḏawla* dialects, or an orientation to that speech community, for the allocation to have gone in this direction.

It seems likely that this *ḏayla*-feminine speech community formed only once, since the distribution of plural forms could certainly have gone in another direction in a different circumstances of contact. Geographically, a good candidate for the location of its formation is in Northwest Yemen (or somewhat further north, due to possible interposition of OSA speaking groups), where dialects with *ḏayla* as a common plural occupy the coasts, but the highlands have *ḏawla* dialects which transition to *hawla* dialects, and most of the dialects with feminine plural *ḏayla* are sandwiched between these two groups. The feminine plural *ḏayla* group probably was somewhat more widespread than it is today, given the location of a isolated member of this group on the coast (Yemen #84) where it appears to be getting displaced by other dialects, and the importance of these groups to the history of Arabic.

The historical data also supports an interpretation of this dialect group aris-

ing in or near the Yemeni Tihama. As discussed in Section 4.1, the initial colonizing presence in Egypt is said to have come from speakers of Yemeni Tihama dialects, with a very large portion of the settlers coming from the relatively small tribe of ‘Akk. According to Caskel (*EI2*, “*Akk*”), in the seventh century this tribe’s southern territory extended as far as al-Ḥudayda, which is Yemen #84, the only surviving coastal dialect with feminine plural *ḏayla*.

This was a relatively small group of ‘Akk, reportedly some four thousand troops in the initial invasion, soon dwarfed by later contingents that increased the population of the invading army to some twenty thousand troops. However, many of the later troops may also have been from the same region. This illustrates the power of the principle of first effective settlement: an initial colonizing population of some fourth thousand or maybe at a stretch twenty thousand speakers drawn from this speech community have been able to establish a dialect that is now used by over eighty million speakers.

There is one difficulty here, which is that the current dialect in Yemen #84 may be a consonant-alternating type, with demonstratives *ḏa:*, *ta:*, *ḏawli*, *ḏayla*, though the maps are not clear and the singular forms may be marking its nearest neighbor, Yemen #85. The general trend for *ḏawla*-type dialects to be vowel-alternating, not consonant-alternating, suggests that Yemen #84 was originally a vowel-alternating dialect, like Nile Valley dialects today. Almost all of the surrounding dialects today are both consonant-alternating, and have only common plural *ḏayla*, as expected, and may represent a later layer of settlement. The surrounding countryside would have shifted to the incoming dialect, but the higher population density of a city like al-Ḥudayda would have preserved the earlier dialect forms.

Another issue is the lack of a feminine demonstrative in the modern dialect

of Cairo (EGCAI). Almost all of the other Nile Valley dialects (EGBER, CHDSH, SHUK, NIG), on the other hand, show variants of a feminine plural *de:l*. We can therefore explain the loss of the feminine plural in Cairene Arabic as a result of the general loss of feminine plural demonstrative forms that characterizes most major cities in the eastern Mediterranean. This loss is clear in both the major urban areas of Syria, which would almost certainly have been part of the areal speech community which included Cairo. In Baghdad, not only do we seem to have a historical loss of feminine plurals, but there is even an ongoing loss of feminine plural forms from rural dialects once they move into the city, as discussed in the following section. This seems to be the result an urban-hierarchical-type diffusion of the loss of these forms, with the change affecting major cities, and their nearby satellites, but often leaving places in between untouched.

CHD has no feminine plural forms reported in Hagège (1973), but Roth (1979, p. 169) records it as a marginal form, and the nearby dialect of CHDSH also shows *de:l* as a feminine plural. Here we have a dialect which is currently in the process of losing the feminine plural form, which is not entirely unexpected given its very marginal use (only for groups consisting entirely of female humans).

This suggests that the primary colonizing dialects of at least Fustāt and by extension Cairo was of the *de:l*-feminine plural type. Though today, the dialects of Upper Egypt, Nigeria and Chad are of the same type as Cairene, we have evidence (discussed in Section 6.4 below) that Upper Egyptian dialects were originally of a different type. This suggests that the Cairene-type demonstratives eventually replaced those original dialects, a process that was largely completed by the fourteenth century, when Arabic speakers began moving further south into Sudan and then Chad and Nigeria.

The Benghazi NABEN dialects, which have forms *'haḏa~'haḏʿa*, *haḏi*, *ha'ḏʿo:l*,

ha'ðe:n might in fact reflect the *ðe:l*-type dialect, if we understand the final *-n* to reflect a similar ad-hoc change of /l/ > /n/ that also occurred in dialects with *ðanna*. This change would have been helped by the association of /n/ with feminine plural forms, or could have been a result of contact with the *ha:ðu:m* that migrated into North Africa. Benghazi is only a short distance from the city of Barqa, which was conquered and colonized by the same forces which had conquered Egypt, which favors an interpretation of these endings as reflecting original Nile Valley-type dialects. The **ha:-*prefixes in the demonstratives, not found in Nile Valley dialects, could be the result of subsequent contact with other North African dialects, which Benghazi Arabic also resembles in many linguistic features.

W. Fischer (1959, p. 114) suggests an alternate explanation for the Benghazi forms, reconstructing *ha'ðe:n* as *ha:ða:-hen* i.e. a form of presentative preceding the feminine plural pronoun. But while there are more elaborate pronouns developed along these lines in this dialect (e.g. *ha:ði:kka:hen* 'that (feminine?)'), this does not seem to be the general strategy in the dialect. More likely it was formed in analogy with the pronouns, but the possibility of an /l/ > /n/ shift from an original **ðe:l* remains.

6.3.3 *ha:ðo:l* and *ha:ðo:l*, *ha:ðanna* dialects

Another subset of dialects with masculine plural *ðawla* innovated the unusual feminine plural *ðanna*, derived from *ðalla*. This unusual change, from /l/ to /n/, defies a clear explanation at this time (see Section 5.3.2.2), but it is so unusual that it makes an excellent diagnostic, since it probably only occurred once. These dialects originally did not have **ha:-*prefixed demonstratives, and later developed them in-situ, but because most of the modern descendants have **ha:-*prefixes, at least optionally, I will refer to these as *haðanna* dialects.

One might consider there to be two types of these dialects, those with just *ha:ðo:l*-type demonstratives, and those which have *haðanna* also. However, it seems likely that in most places, these *haðo:l* dialects actually reflect an earlier *ha:ðo:l*, *haðanna*-type dialect, with subsequent loss of the feminine plural, and so we will treat them as deriving from a single dialect group pending more information, especially on the dialect geography of modern Saudi Arabia.

The area where we can most clearly see the history of these dialects is in Syria, where we have extremely detailed dialect maps. In the Syrian dialects today (see Figure 6.2) we find the *hawla*-type dialects between the coast, and the urban centers, which essentially form a north-south line from Aleppo to Damascus. To the east of these cities, almost all of the sparsely populated steppe areas, and the small villages around the Euphrates have demonstrative plurals *ha:ðo:l*, *haðanna*. In the cities themselves, we have only *haðo:l*, but the prevalence of forms with the feminine plural outside of the cities suggests that these dialects have simply lost the feminine plural.

A similar change may have occurred in Iraq. Here, the archaic city dialects (i.e. Christian and Jewish dialects) also have *haðo:l*, as do the more recent Muslim dialects. However, these Muslim dialects still have an optional feminine plural (*ha*)*ðanni*, which seems to be, like the feminine plural pronouns, more characteristic of rural *gəlat* dialects. Indeed, in a note about the feminine plural pronouns, Blanc (1964, p. 60) remarks:

As for the masculine-feminine distinction in the 2pl and 3pl which is a hallmark of Beduin dialects, the rural *gəlat*-dialects have forms like *entu* and *humma* for the masculine vs. such forms as *entan* and *henna* for the feminine; my M[uslim] informants have heard such feminines, but do not use them and characterize them as provincial.

Hence, loss of feminine plural forms does appear to be characteristic of urban areas, and when a dialect urbanizes, it tends to lose those forms, as the urban dialects in Egypt, Syria, and Iraq. This spread seems to be urban-hierarchical in nature, with urban areas like Damascus, Cairo, and Baghdad losing their feminine plural demonstratives. This diffusion is not of linguistic material per se; rather what is spreading is the understanding of what categories language should mark, the kind of diffusion behind metatypic change, seen very clearly in the spread of evidentiality marking (see Section 1.3.3.4). This understanding transfers between speakers of the larger urban speech community, since they have stronger network ties with one another than they do with those living in the near countryside, and probably also because they see themselves as urban dwellers who should act as such and not appear “provincial.”

Almost all of these *(ha:)ðo:l*, *(ha:)ðanna* dialects have **ha:-*prefixed demonstratives, but this does not seem to be a feature of these dialects originally, but rather a later series of alignment events. While most of the dialects in Syria have **ha:-*prefixed demonstratives throughout their paradigms, we can find occasional holdouts. Syria #361, just north-east of Damascus has the demonstratives *ða:~(ha:)ðe*, *ði~ha:ðe*, *ðawl* | *ðɑ:k*, *ði:k*, *ðoli:k*, with optional **ha:-*prefixed demonstratives only in the near proximal forms. This dialect is close enough to Damascus that it might have lost feminine plural due to urban influence, so it might have had *ðanna* as well at some point in the past.

The dialects of LVSPAL and LVSSOU, Palmyra and Soukhne respectively also seem resistant to the spread of **ha:-*prefixed demonstratives, but are adding them nonetheless. Each dialect is diffusing the **ha:-*prefix throughout their paradigms differently, in Palmyra in the proximal demonstratives first, while in Soukhne they are obligatory through the singular demonstratives, both proximal and distal. The

difference in realizations strongly suggest that each dialect has gained the **ha*-prefixes through diffusion, participating in a Syrian urban hierarchy where **ha*-prefixes are largely obligatory, but each dialect conforming to expectation in different ways.

The *ḏanna*-dialects probably did not arise as far south as Yemen, where we only have attestations of dialects with *ḏawla*, *ḏalla*, and we don't see them further north in Al-Qauz (see the demonstratives for ARQz). Meccan Arabic, which Bruce Ingham (1971) notes differs significantly from other Hijazi dialects today, has proximal plural *hado:l* but no feminine plural (Abu-Mansour, *EALL* : “*Meccan Arabic*”). According to Ingham, Mecca participated in the speech community of Damascus and Cairo and thus it could have lost the feminine plural through the influence of those dialects.

It is difficult to posit when these dialect arose; nor does knowing their birth date tell us how or when they became important dialects. Of course, the *ḏalla*-dialects should have been produced as part of the same process that produced *ḏayla*-dialects, i.e. around the second century CE, though again this dating rests on the tradition that Omani dialects date to a migration in the second century. There is no way to know when exactly the shift from *ḏalla* > *ḏanna* occurred, however, especially given how poor our current knowledge of Peninsular dialect geography is.

In some ways, the ultimate time of origin of these dialects is beside the point, as what really matters is when these dialects became prominent and widespread. They certainly seem to have risen to prominence much later than the other dialects discussed here — in the Levant and Iraq, *ḏanna*-dialects are all closer to the steppe than the other dialect types, and this implies a later arrival of those dialects.

A possible scenario for how *ḏanna*-dialects came to prominence is that they were the original dialects of the *ʿamṣār* in Iraq. Perhaps they were survived, or developed from the koineization process in Kufa, which went on to become the dialect of Baghdad, where only the form **haḏo:l* survives in the generally conservative IqBGJ.

There are several reasons to support such a scenario. First, these dialects are spread most successfully along the Euphrates, while further east and along the Tigris, remnants of *hawla* have been better preserved. Kufa's location, west of the Euphrates, would support this scenario, with later diffusion out of Baghdad filling in the areas along the Tigris. The diffusion of *ḏanna*-dialects along the Euphrates does suggest an upriver movement, and from there it may have moved into the cities of the Levant, with a large number of *ḏanna*-dialects attested in the areas around Aleppo.

Second, in Basra, while we have clear traces of *ḏanna* dialects in the proximal demonstratives ((*ha*)*ḏo:l*, (*ha*)*ḏanni*), the distal feminine plural demonstratives appear to show forms which suggest a calquing-type change, e.g. *ḏa:kan* and *ḏi:kan*, where feminine plural demonstratives did not exist, but were produced by adding feminine plural suffix pronouns to the singular demonstratives (Mahdi, 1985, p. 155). However, these forms shows two things: first, they show that the *ḏanna* type forms are relatively new to this dialect, since distal demonstratives tend to preserve older forms. Second, they show that speakers of this dialect felt there was a need for feminine plural demonstratives, but that they had not yet borrowed the *ḏanna*-type forms; if they had, then the distal feminine plural form would be *ḏannak* (which is attested, but only one of many forms.) These pieces of evidence suggest a scenario where Basran Arabic had its own, distinct demonstrative forms, but came increasingly under the sway of *ḏanna*-dialects. First, there

was an attempt to maintain the original demonstrative forms, but with new forms calqued to match the expected category of feminine plural demonstrative; after this time, a general shift occurred, with demonstratives borrowed wholesale from the *ḏanna*-dialects

In Mosul, we have distal plural demonstratives of the form *haḏo:k* (no feminine plural), which is characteristic of a Stage 2 *hawla* dialect shifting to a *haḏo:l* dialect. Mosul was built and colonized by speakers from Kufa, but was built near an existing settlement. Near to Mosul today, but on the other side of the Tigris, we have a settlement called Bəḥzāni whose demonstratives are a textbook example of a Stage 5 *hawla* dialect. This suggests that at around the time of the foundation of Mosul in 641 CE, the area was likely already speaking a Stage 2 *hawla* dialect, and the Bəḥzāni dialect maybe have developed later to Stage 5. Of course, the original settlement of Mosul was too soon after the settlement of Kufa in 638CE for a koineization process to have been completed, but we have accounts of later settlement, and the *haḏo:l* forms may have arrived via hierarchical diffusion from Kufa/Baghdad as they probably did in Basra also.

Indeed, along the upper Euphrates generally, we find many of these shifted Stage 2 dialects, as discussed above in Section 6.2.3, and shown in Figure 6.3. The city of Hīt, not that far north of Baghdad, shows *ha:ḏo:k*-type dialects, and it seems likely that all of the dialects upriver of at least Hīt were of the same type. As discussed in Section 6.2.3, the *ha:ḏo:k*-dialects are the result of a type of calqueing process, which implies that the Stage 2 dialects were well entrenched before the *ḏanna*-dialects became more prestigious.

This scenario matches a Kufan origin of the *ḏanna*-dialects quite well: The Stage 2 dialects would have been the original dialects spoken in Iraq, particularly upper Mesopotamia where the Arabic presence was better established in pre-

Islamic times. The Kufan dialect was established with the creation of the city, and probably stabilized shortly thereafter. As Kufa and then Baghdad became the economic centers of Iraq, the speakers of the original Stage 2 *hawla*-dialects would have been under pressure to adopt the prestigious central Iraqi dialect forms. At the same time, these dialects had a long history and probably local solidarity, and so they accommodated partly to the prestigious dialects, while still marking their own dialects as distinct. Later on, some of these dialects shifted more completely, as in the case of Hīt and Al-Bukmal.

As we stated in Section 3.6.4, the Arabization of the Iraqi countryside was probably driven by the nearest major city, that is Kufa/Baghdad in the north, and Basra in the south. This would have created a large number of speakers of *ḏanna*-type dialects, which would explain why these dialects are found in the Iraqi countryside also.

It is not clear what led to the development of *ḏo:la*, *ḏanna* dialects in the cities of the Levant. The dialect geography seems to suggest a movement from further east, possibly from the dialects which have moved up the Euphrates. As mentioned above, Aleppo could have been one of the first cities which acquired this type of demonstratives, which then moved south towards Damascus. This probably would have happened after the end of the plagues at earthquakes in the Levant in 749CE. The movement of power to Kufa in 750 CE, and Baghdad in 762 CE as a result of the Abbasid revolution would also have increased the likelihood of hierarchical diffusion from the newly important capital into the now-secondary (and possibly still poorly-Arabized) Levantine cities.

In the Levant too we have the familiar pattern of *ha:ḏo:k* dialects which seem to be replaced later by *ha:ḏo:l* dialects. The *ha:ḏo:k* dialects could have begun diffusing directly from Iraq into the areas around or in the major cities, and then

only much later was the pressure of *ha:ðo:l* enough to full shift the urban dialects to the use of *ha:ðo:l*.

There is also the likelihood that *ðanna* and certainly *ðawla* type dialects continued to move throughout the Arabian peninsula, and thus the picture, especially in the Levant, is rather complex. Many of the dialects we consider to be of the modern “Bedouin” type have *haðo:l*-type demonstratives, though not always with *haðanna* (but see Rosenhouse, 1984, p. 20). In ARNJD we have a feminine plural form clearly derived from the masculine plural *(ha)ðo:la*, *(ha)ðo:li* using the usual *-i* feminine formative. There is no way to know whether dialects such as this ever had *ðanna* as their feminine plural, so we must remain agnostic on this point for now.

One speculative possibility is that many of the modern Bedouin dialects in the Levant and Iraq actually derive their demonstratives from these southern Iraqi *(ha:)ðo:l*, *(ha:)ðanna* urban dialects, which diffused first to the rural dialects, and then possibly to Bedouin dialects. In the ninth and tenth centuries, the Abbasid empire faced a series of crises, many of which adversely affected the countryside. These crises may have forced rural dwellers to rely less on immovable property and more on movable livestock, switching from a fully agricultural to a small-stock, semi-pastoral lifestyle (Waines, 1977, p. 303, esp. fn. 65). In this case, it is possible that the modern (esp. Syrian) Bedouin *ha:ðo:l*, *ha:ðanna* dialects are ultimately derived from early Islamic urban dialects in Iraq, by way of rural Iraqi dialects.

To summarize, Northern Iraq was probably colonized by *hawla* dialects in the pre-Islamic era, probably by the fourth century. After this, the koineization process in dialect Kufa resulted in a final dialect (or at least, a prestige dialect) with *ðo:l*, *ðanna*-type forms. These forms diffused to the other large cities of this

region, Mosul and Basra, where they changed and in Basra, eventually displaced, the local native demonstrative forms. These *ḍanna* type forms diffused up the Euphrates, where they produced the *ha:ḍo:k* inter-dialect form. It is possible that the *ḍo:l*, *ḍanna* moved into directly into Aleppo, and then other Levantine cities, though there might have been diffusion directly from Kufa and then Baghdad into cities like Damascus and Hama, a process that probably started after the end of the plagues and earthquakes and at the beginning of the Abbasid era when Iraq became the seat of power. A series of crises in Abbasid Iraq, beginning in the ninth and tenth centuries, may have prompted settled peasants using *ḍo:l*, *ḍanna* to become partly nomadic, and this may be the ultimate source of modern *ḍo:l*, *ḍanna*-type nomadic Bedouin dialects. Indeed, those nomadic dialects may also have played a role in importing the *ḍo:l*, *ḍanna*-forms into Syria.

However, it is important to hold in mind that this complex scenario relies heavily on the limited data from Iraq, much of which has been gathered by speakers living outside of the country and their normal sociolinguistic context. Until we have better dialect data on Iraq and the rest of the Arabian Peninsula, it will have to remain a very speculative account.

6.4 Dialects with *ha:ḍu:* plurals

The speech community which is in some ways most mysterious is that of the north African dialects which share the plural form *(ha:)ḍu:*. This form is not attested as such in the data available except in North African dialects (and some African) dialects, and as discussed in Section 5.3.2.2, cannot be ascribed a relative chronology in comparison with the other dialect groups.

The closest dialects attested in the Arabian Peninsula are those in southern Yemen which have clearly derived their plural forms from pronominal suffixes, as

in Yemen #158 with plural proximal *ʔa:ðum/ʔa:ðin*, and we do find similar forms in Tunisia, though W. Fischer (1959, p. 113) claims the North African forms with nasals developed in-situ from forms such as *ha:ða:+humma* > *ha:ðum*, but it is necessary for him to assume a number of complex phonological changes. It is more likely that both the Yemeni forms and North African forms are derived directly by analogy with the endings of the plural pronouns (both independent and suffixed), which in many dialects are masculine *-um* and feminine *-in*, though some dialects have levelled these endings further.

It is possible that *ha:ðu:* formed from original *ha:ðum* but lost the nasal from the paradigmatic pressure of the other vowel final demonstrative forms (e.g. *ha:ða:*). NATrM dialects might show such a transitional dialect, where a masculine plural form *ha:du:ma* is attested along with *ha:du*, but the distal forms all derived from *ha:du*, so those forms are themselves probably original, with the nasal forms reflecting an later set of migrations.

The nasal-final dialects do appear to be the result of a later migration than the one that brought *ha:ðu:* dialects to North Africa. The nasal-final dialects are primarily attested further east, only spreading as far as Taza in north-east Morocco according to W. Fischer (*ibid.*, p. 113), that is to say, not spreading past the Middle Atlas mountains, where we only find **ha:ðu:*-dialects. In places that do have the *ha:ðu:m*-type demonstratives, they almost always still show **ha:ðu:*-type demonstratives in the distal, further evidence that the nasal-final dialects are a later arrival.

Even Mauritanian Hassaniya dialects (NAHAS) has *ha:ðu:*-type demonstratives, though it is commonly considered to be a direct descendant of the Hilalian dialects. Hassaniya dialects are said to have arrived as a subgroup of Banū Maʿqil as part of the Hilalian invasions, and were resident in Morocco starting in the twelfth

century, and moved into the Western Sahara and Mauritania in the fourteenth and fifteenth centuries (Pellat, *EI2*, “*al-Maʿkil*”). It is therefore possible that these dialects adopted the mainstream Moroccan demonstratives while resident there, before bringing them further west. In that case, we could designate the *ha:ðu:* dialects as the original pre-Hilalian dialects, while the *ha:ðu:m* dialects would neatly correspond to the Hilalian dialects.

Instead, as discussed at length in Section 4.2.1, most of North Africa was probably Arabized well after the Hilalian invasions, and those invasions were probably simply part of a larger, longer series of migrations westward. The dialect of Cherchell (NACHR) is very clearly of the *ha:ðu:*-type, but it was basically uninhabited based on our sources until *after* the Hilalian invasion, and is reported to have been closely integrated with pastoral society in the twelfth century, which were at that time almost certainly composed of Hilalian tribes. The same is true for the other major ‘Pre-Hilalian’ dialects in our sample, NADJ and NATL.

The conformity of *NaHas* with general Moroccan norms is also surprising given that dialect’s divergence from most North African norms, as noted by Taine-Cheikh (*EALL* : “*Ḥassāniyya Arabic*”):

Despite the influence of the [Berber] substrate and because of its Bedouin nature, it has more in common with eastern Arabic dialects [...] than with most of the Maghrebi dialects like Moroccan and Algerian.

If this dialect is relatively uninfluenced by North African dialects, it seems amazing that it would have adopted the earlier North African demonstrative forms.

Thus, it seems likely that the *ha:ðu:* dialects may actually represent a relatively late set of dialects, moving into North Africa as late as the twelfth century,

while the *ha:ðum*-type dialects simply represent a different set of migrations which came later, but still significantly impacted the dialects of North Africa.

We may not be able to actually recover the early Arabic dialects in North Africa, at least not in terms of their demonstrative forms. In Tunisia, we have do find adnominal demonstratives of the type *ha:*, *ha:k* instead of the more general North African *ha:d-*, which could point to an underlying *hawla*-Stage 1 dialect, though it could also be an independent innovation following erosion of *ha:d-* > *ha:*, as argued by W. Fischer (1959, pp. 44-46). Tunis was said to have been founded by a commander from Syria presumably with Syrian troops, so it is possible that the Tunisian dialect does preserve an earlier form imported from the Levant.

Tunisia, however, has quite a few different demonstrative forms, representing many different layers of immigration. Tunisia was often the ultimate destination (whether intended or not) of western migrations. The relatively flat, fertile land would probably be more appealing than the narrow coastal littorals in Libya and Algeria, while the way further west to the flatter lands of Morocco would require crossing the Middle Atlas. As a result, we have a bewildering diversity of demonstratives in Tunisia: We have *ha:ðu-*-type demonstratives, *ha:ðu:m*-type, we even find demonstratives of the type **ha:ðu:la*, possibly deriving from *ha:ðo:l*-type dialects. The city of Susa even has *hɛ:ka*, *hɛ:ki*, *hɛ:kum* for one of its distal demonstrative series, which is most similar to the dialect of Sason in Anatolia (Talmoudi, 1980). A thorough dialect survey of Tunisia might find some remnants of earlier North Africa dialects — the antiquity of such remnants could be established by comparison with e.g. Maltese, which at present is an outlier.

The dialect situation in North Africa represents a far more complex mosaic of migrations and mixing than can be accounted for by the simplistic pre-Hilalian/Hilalian dichotomy. As argued in Section 4.2.1, the non-linguistic histor-

ical records do not support this dichotomy. Here, the linguistic data, even just for the demonstratives, provides further evidence that multiple migrations, at different times, are responsible for the dialect geography of North Africa. Some of these migrations appear to have occurred around the eleventh century, but they correspond better to the supposedly ‘Pre-Hilalian’ dialects than to the Hilalian dialects.

Tracing the ultimate place of origin of the *ha:ðu:* dialects is difficult. As mentioned above, there are no dialects in the Arabian Peninsula which shows this type of dialect. One of the few places outside of North Africa where we find a *hadu* form is in the Sudan, where it is used as a presentative (W. Fischer, 1959, p. 177). Indeed, the form of the upper Egyptian “distal” forms, which are actually clearly derived from presentatives, look like they must have been derived from *ða:*, *ði:*, *ðu:*-type dialects, since they show the very same vowel alternation, just with the /-k/ distal marker. NAHAS has presentatives very similar to Egyptian dialects, with Taine-Cheikh (*EALL : “Ḥassāniyya Arabic”*) giving *ðik-hi(yyæʾ) mirʿyæm* (this.FEM-her miriam) ‘This is Miriam.’ Similar presentatives are found in Hijazi Saudi Arabian dialects, which might preserve a trace of the parent dialects of North African *ha:ðu:*-dialects: Theodore. Procházka Jr. (1988, p. 226) records a form *duk* ‘look here’, which can be optionally conjugated based on the gender of the interlocutor (and not on the gender of the indicated object) as *duk*, *dutʃ*, *dukum*, *dukin*, ‘f’or masculine, feminine, masculine plural and feminine plural respectively.

We also have evidence that the shortened *had-* adnominal demonstrative, found in most North African dialects, was current as early as the third hijra century, when it is found used in papyri (Hopkins, 1984, p. 66).¹⁹ The ancestors of North African type dialects were therefore clearly in Egypt by that time. This would date

¹⁹ These papyri are most likely from Egypt and the Nile Valley, though I have not yet been able to trace all the references to be sure of this.

the existence of such forms in Egypt to after the arrival of the so-called ‘Qaysiyya’ to Egypt in 109 AH/727 CE, which included elements of the Banū Sulaym and Banū Hilāl tribes (see Section 4.2.1), and so we can imagine that these were the carriers of the dialects which have *ha:ða:*, *ha:ði:*, *ha:ðu:*-type demonstratives, and which, while in Egypt, developed shortened adnominal forms of these demonstratives.

The following scenario thus suggests itself: As early as 109 AH/727 CE, a group of speakers from various tribes in the Qays confederacy moved into Egypt, speaking *(ha:)ða:*, *(ha:)ði:*, *(ha:)ðu:*-type dialects. In the areas around Fuṣṭāṭ speakers were predominantly of the *ða:*, *ði:*, *ðu:l*, *ðu:l* dialect type. As noted previously, these dialects likely had at least optional **ha:-*prefixed demonstratives in the proximal demonstratives, but probably did not have **ha:-*prefixes in the distal demonstratives, hence we have the grammaticalized adnominals *ha:d-* (proximal) but *ðVk-* (distal) not *ha:ðVk-*.

During their time in Egypt, those dialects began developing their adnominal demonstratives, at least of the *had--*type, possibly replacing the previous *ha:-*. Following this, these dialects expanded into North Africa, and over time, their demonstrative forms may have replaced the demonstrative forms which were current in the area, or Arabic may simply not have been as well established as we have previously believed. When exactly this migration happened, we cannot be sure, but it is entirely possible that it was part of the historiographical Banū Hilāl migrations, since as discussed here and in Section 4.2.1.3, many of the supposedly “Pre-Hilalian” towns were actually probably founded or Arabized much later, at times which post-date the mythical Hilalian invasions.

The traces of these dialects have largely been erased from Egypt today, but persist in the presentatives-cum-distal demonstratives, particularly those used in Upper Egypt. The fact that demonstratives generally are a site of intense soci-

olinguistic evaluation in Egypt means that such a scenario is not unexpected — the dominant variety, that of Cairo, would be able to impose its (proximal) forms over the rest of the country through diffusion. Moreover, the fact that there are marginal attestations for *hadu* in Sudanese dialects suggests that there was still some traces of these forms when the ancestors of the Sudanese dialects moved out of Egypt into Sudan as early as the twelfth century CE. This suggests that these dialects were dominant in Upper Egypt from about the eighth to the twelfth century, some four hundred years.

The dialects which have final nasal forms are more recent, which is supported by the fact that they largely occur in the proximal sets, while the distal demonstratives are more clearly derived from the original *ḏa*ː, *ḏi*ː, *ḏu*ː-type dialects. Moreover, these dialects are all among the more easterly dialects (e.g. Tripoli and Tunis), again supporting the notion that they derive from a migrations coming from the east into North Africa. Their ultimate origins are also unclear. There is a set of similar dialects in Southwest Yemen, but they are surrounded by difficult terrain and different dialects on all sides and seem to be present in largely agricultural areas, so it seems unlikely that speakers of these dialects could have moved out into North Africa from their current Yemeni location. They could, however, share a common ancestor, though the development of these nasal-final demonstratives is logical (direct analogy with the pronouns) and therefore they could have developed in different places at different times.

6.5 Maltese Arabic

Maltese Arabic has apparently quite different demonstratives from those of the mainland North African dialects. Modern dialects have the following primary forms: *da*, *di*, *daw* | *dak(a)*, *dik(a)*, *dawk(a)*, but the proximal forms also have

variants *dan(a)*, *din(a)*, *dawn(a)*, and historically there were also forms with **ha:-* prefixes, *heda:na*, *hedi:na*, *hedauna*, though these are primarily attested in literary registers (W. Fischer, 1959, pp. 67-71).

We must first account for the plural forms, with or without final *-n(a)*. Fischer suggests that this form developed in analogy with the final weak verb third persona plural verbs conjugations with underlying *-a:*, illustrating these with the forms *ar-au* ‘You (pl.), look!’, to which I will add *idra* ‘You (sg.) get used to’ vs. *idraw* ‘You (pl.) get used to’. In the imperfect plural, this class of final weak verbs also shows *-ew* or *-aw*: *nidraw*, *tidraw*, *yidraw* ‘we know, you (pl.) know, they know’.

This seems an obscure source for such a change, but these final weak verbs are an important part of Maltese: First, they absorbed other verb classes, so for example older *fasa:*, *yafsu:* ‘to break wind’ was integrated into the final imperfect *-a:* verb class as *fesa*, *yifsa* in Maltese, and the merger of /ʎ/ and /ɣ/ > /ʎ/ > /Ø/ led to verbs with those final consonants also being treated as final-weak, though often in the /i:/ final class. Then, as this class grew increasingly important, the final weak conjugation was extended to include some Romance loan-verbs, so one finds *kantaw*, *ykantaw* ‘They sang, they sing’ < Romance *cantare* (Mifsud, 1995, Chapter 4, Appendix A–C).

The problem is that it is not clear why the suffix *-ew* wasn’t generalized instead of *-aw*, and how exactly this form was taken from the verbs. Most of the demonstrative systems here seem to be primarily based on the pronouns, with further analogy made possible by verbal affixes. We do find some *-aw* forms in the pronominal system — in both dialects (St. Juian and Marsaxlokk) surveyed in Schabert (1976), the third person masculine possessive pronoun is *ti:aw*. Marsaxlokk generally merged the /u/ with /æw/: this produces the independent masculine sin-

gular pronoun $\text{æw}(w\text{æ})$ (< $*hu(wa)$) and plural æwmæ , and all plural forms verb forms have /-aw/ suffixes. Both dialects have *daww* for their plural demonstratives, however, and Marsaxlokk does not have the expected *dæww* if it were raised from $*du$.

Furthermore, Schabert (1976, p. 67) lists as alternative plural forms, *da*, for both dialects, which is quite unexpected. None of the dialects surveyed here should have a form like that in the plural, and so it is difficult to classify it.

With regards to the development of the final -n forms, W. Fischer (1959, p. 70) argues that these developed due a tendency for -n to be lost word finally. In a number of words, we find doublets, so we have *fein~fei* ‘where’ and *aun~au* ‘here’, which Fischer believes would have been transferred to the plural demonstrative *daw* as speaker misinterpreted it as being originally *dawn*, and it would have moved from there to the rest of the demonstrative series. This explanation is not entirely satisfying, but I see no clear alternative²⁰, and so it would seem that the development of final -n in the demonstratives is a later development in Maltese, and probably not characteristic of the dialects which it is derived from.

The unusual nature of the Maltese demonstratives may preserve the original demonstratives in North Africa which existed in Tunisia prior to the later population movements into North Africa in the tenth and eleventh centuries. Malta is said to have been colonized largely from Sicily, and Sicilian Arabic is probably originally a product of ninth century invasions which departed from Tunisia. However, these demonstratives are extremely unusual compared to most other Arabic dialects, and the development of final nasals seems to be a relatively recent feature. Therefore, we cannot rule out the possibility that Maltese has innovated

²⁰ Recall that Punic, which had a masculine singular demonstrative *zn* is no longer believed to have formed a substrate in Maltese. See Section 4.4.1.

this demonstratives separately from other dialects, so that they do not reflect the archaic early North African dialects.

6.6 Classical Arabic

In looking at the Classical Arabic demonstratives, the following description seems to fit:

The [...] creation of written languages may be compared to the formation of a film of ice on the surface of the river. The ice borrows its substance from the river [...] and yet it is not the river. - Joseph Vendryès, trans. Paul Radin, cited in Joseph and Janda (2003)[144].

The demonstratives in Classical Arabic, especially in Quranic Arabic, seem to reflect the shifts in speech communities which probably passed through the Western Arabian peninsula: The more grammaticalized relatives markers reflect a straightforward consonant-alternating dialect with initial *ʔalla-* prefixed to *ði*, *ti*, *ði:na*. The distal demonstratives (*ða:lika*, *tilka*, *ʔula:ʔika*) appear to reflect a consonant-alternating dialect also, though they have some other changes. The proximal demonstratives appear to reflect the demonstratives *hawla* (*ha:ða*, *ha:ðihi*, *haʔula:ʔi*) speech community, whose sociolinguistic prominence came after that of the consonant-alternating dialects in the northwest Arabian peninsula and Levant.

The proximal feminine form *ha:ðihi* is likely a generalized pausal form of *ha:ði*, a phenomenon widespread in dialects throughout Yemen (e.g. points #9, 14, 18, 59, 130 among many others), including *hawla* dialects. The distal form *haʔula:ʔi* may simply reflect an addition of glottal stops that were not original in the dialect it was drawn from. The original orthography in the Quran, and that of many papyri

(Hopkins, 1984, p. 67) normally is closer to the dialectal form without glottal stops, reflecting *hawlay* or *ha:wla:*. al-Mubarrad (*al-Muqtaḍab*, pp. 278-9) notes that these forms may be pronounced without the glottal stops, but it is preferable to include them.²¹

The distal demonstratives, *ḏa:lik*, *tilka*, *ʔula:ʔika*, are unusual. On the one hand, they are clearly consonant-alternating dialects, which fits with the “layering” process that seems to have happened in this register. On the other hand, the -*l*- element is extremely difficult to account for, and as discussed previously, no etymology has been proposed for it that I find convincing. It is not attested in any modern dialect, which makes it difficult to develop a clearer picture of how these forms relate to the dialects from which Classical Arabic may have been formed. Moreover, these forms also seem to show a kind of transition to vowel-alternating patterns — the masculine has /a/ vowel in the first syllable, and the feminine has an /i/ vowel. This could be contamination from the singular, or record a transitional period that occurred in the dialect that gave rise to these forms. Recall, however, that there are no attestations in modern dialects of such mixed forms, though this may be a result of our limited data on Saudi Arabian dialects.

This is not to say that the Quranic Arabic forms are necessarily ‘artificial.’ It is not entirely unusual to find dialects which seem to preserve multiple layers of forms. A number of dialects in southern Yemen have vowel-alternating demonstratives in the proximal, but consonant-alternating demonstratives in the distal (see Table 5.7 for representative forms). Similarly, the dialect of Yafi^c reported by

²¹ One wonders whether the insertion of glottal stops in the Quranic recitation forms is a result of the same process in Iraqi Arabic where a glottal stop is inserted between two vowels that would otherwise form a diphthong. Such a dialect would potentially borrow a form like *hawlay* (as attested in some dialects with a final diphthong) and make it into *haʔulaʔi*. See the discussion of this in Section 5.4.4.

Vanhove (2004) shows primarily vowel-alternating demonstratives now, but has fossilized and grammaticalized uses of **ta:*. So it is possible that Quranic Arabic actually does reflect a local dialect, at least in terms of its demonstrative system. If Quranic Arabic really reflects Meccan usage, we might even expect to find layering in that dialect due to the numerous migrations that passed through the central Hijaz.

On the other hand, the preservation (or innovation?) of the dual demonstratives, and the bizarre -l- medial distal demonstratives, both of which are apparently absent from all other dialects, suggest that the situation is not nearly so straightforward. Certainly Quranic Arabic is a rhetorical, elevated register, and it shows numerous archaisms throughout its grammar. It is very likely that the demonstrative forms characteristic of Quranic Arabic represent a relatively archaic local dialect preserved in a literary register, not the dialect that was in use at the time of the prophet.

Outside of the actual usage of the Quran, the Arabic grammatical tradition reports many more demonstrative forms. Some of these forms are given in poetic citations, but many are not, and so it is not entirely clear what sources the grammarians gathered them from. We have the following forms reported by Sībawayhi (1999, vol II, p.75):

Demonstratives include (*al-ʿasmāʾ al-mubhama*): HDA:, HDA:N, HDH, HA:TA:N, HWLA:ʔ, DLK, DA:NK, TA:NK, TYK, ʔWLYK

And later on the same page he discusses a form DA:K which is of course more widely attested in modern dialects than DLK. al-Mubarrad (*al-Muqtaḍab*, p. 275) actually treats DA:K this as the more basic form, with a “meaningless (*zāʾid*)

l” that can be added to create *ḍalika*. al-Mubarrad (*al-Muqtaḍab*, p. 277) also gives further feminine forms: *ḍH*, *ṭH*, *ṭA*: all with apparently the same meaning.

These many reported forms do indeed track well with our established dialect geography of Arabic. What is not expected is that there are very few limits placed on these forms, or even the relation of these forms to specific dialects or groups. What attestations we do have of social divisions related to demonstratives are fairly sparse — the grammarians almost seem to treat these as a ‘grab-bag’ of options, with no need for paradigmatic consistency.

Rabin (1951, pp. 152, 203), based on reports attributed to the grammarian ʿAbu ʿUбайд (d. 223/838), claims that the “Tayyi” dialects (and presumably meaning their Classical registers) used the consonant-alternating feminine demonstrative *ta*:. According to Rabin, this tribe resided in the Najd, not the Hijaz. The attribution is too divorced from any larger framework of demonstrative variation to be of great use in understanding the history of Arabic demonstratives.

Similarly, we have reports of fairly minor differences, such as the use of *ha:ḍihi* in non-pausal contexts as an isogloss dividing the Hijaz (where this is the form in all contexts) from the “eastern dialects” where this is only the pausal form, with contextual *ha:ḍi*:.

It is worth noting that, in spite of these grammarians’ reports, the overwhelming linguistic behavior in Classical Arabic texts is to use the “standard” Classical Arabic forms, matching those in the Quran. Searches in arabiCorpus for the non-standard variants find very few examples of these forms outside of the works of grammarians. The primary exception is the form *ḍa:k*, which is almost always used as a discourse demonstrative in the grammatical texts. The one text in the corpus where we do find these forms is Kitāb al-Aghānī, where there is a great deal

of dialogue and use of non-standard forms. Exploring this source further would be of great interest.

Classical Arabic presents, therefore, an unusual situation. Though the Arab grammarians license a whole host of different forms, we almost never find them used in Islamic era texts; instead, the Quranic demonstratives are the primary forms used in literary texts. The many different demonstrative forms reported by the grammarians only really occur (and rarely) in pre-Islamic poetic texts, if we have any attestations at all. The Quranic demonstrative forms do not represent a naturally evolved paradigm — whether by actual dialect contact, or as a result of the process of creating the artistic register, the Quranic demonstratives actually combine many different demonstrative types, representative of different dialects. It is probably not coincidental that the demonstratives it preserves represent important speech communities which, we have argued, passed through the central Hijaz in the pre-Islamic period.

6.7 Dialect Classification

This section places each dialect in the sample into the historical framework developed in the previous chapter, and within the divisions elucidated in this chapter. After each dialect name, the primary proximal and distal forms are listed for the reader to have a better sense of the variation in the order m.sg., f.sg., m.pl., (f. pl.) with a semicolon between proximal and distal forms. Not all forms are attested in the sources for all dialects. Where there are specifically adnominal forms given in the sources, these are indicated; where not indicated, they are not mentioned in the source, and so the pronominal forms perform an adnominal function also.

- **Consonant-alternating dialects**

- ARQAH: *ḍah, tah, ʔawḍi* | *ḍa:k, ta:k, ʔawḍa:k*
- Y96: *ḍa, ta, ḍayl~ḍe:la* | *ḍa:k, ta:k, ḍe:lak*
- YXASH: *ḍa:, ta:, ʔo:ḍa:* | *ḍa:k, ta:k*²²
- Y6: *ḍiyya, tiyya, ʔo:li:, ʔe:li:* | *ḍa:k, ta:k, ʔo:lak, ʔe:lak*
- **Original consonant-alternating dialects with vowel alternation in proximate:**
 - * Y104: *ha:ḍa, ha:ḍi, ḍawla~ḍo:la* | *ha:ḍa:k, ha:ta:k, ḍawla:k~ḍo:la:k*
 - * YTALH: *ḍa:, ḍih, ʔawla:* | *ḍa:k, ta:k*

- **Vowel-alternating dialects**

- **Dialects which had *hawla* as common plural**

- * **No *ha:k* developments**

- AND: *(ʔa)ḍa, (ʔa)ḍi, hawlin(k)~ʔawl(ay)* | *ʔaḍak~ʔik, ʔ, hawlak*²³
 - Y24: *ha:ḍa, ha:ḍi, hawla* | *ha:ḍa:k, ha:ḍi:k, hawla:k~ho:la:k*
 - Y99: *ha:ḍa, ha:ḍi, ʔ*²⁴ | *ha:ḍa:k, ha:ḍi:k, hawlʔa:k*
 - Y121: *ha:ḍa, ha:ḍi, hawla~ha:ḍawla~ha:ḍo:la* | *ha:ḍa:k, ha:ḍi:k, ḍawla:k~ḍo:la:k*²⁵

- * **Stage 2 dialects**

²² The plural form here and in ARQAH is interesting, but without a better sense of northern Yemen and southern Hijazi dialect variation, it's hard to make a clear reconstruction for how these dialects acquired those forms.

²³ The notation used by Corriente (1977) and Zaragoza (2012) is unclear. It appears that there was a original gender distinction in the singular lost as a result of vowel raising.

²⁴ Proximal plural is probably *hawl(ʔ)a*, but might be *haḍawla*.

²⁵ This dialect is obviously difficult to categorize, and lies geographically on the border between dialect groups.

- LvS101: *ha:da, hayye, haw(w) | hada:k, hadi:k, hawki*²⁶

* Stage 3 dialects

- IQAQ: *hada:, hada:~ha(:)dya:, hawda:~hawde: ha:k~hada:k, ha:k~hada:k, hawk~hawda:k*²⁷
- LvS348 *he:ð, ha:y, hawð(i) | hede:k, heði:ki, hawði:ki*²⁸
- LvS239: *ha:da, ha:di, hawdi | ha:ka, ha:ki, hawke:n*²⁹

* Stage 5 dialects

- LvBSH: Pronominal: *ha:da~ha:d~ha:, haydi~hayd~hay, hawdi~hawd~haw | hida:k(i), hidi:k(i)~haydi:k(i), hudi:k(i)~hawdi:k(i)*; Adnominal: *ha- | ha:k-*
- LvS305: *ha:da, he:di, haðo:l | hada:k, hadi:k, hawwi:k*³⁰
- ANDAR: *a:za, ayzi, o:zi | a:k, ayk(i), o:k*

- * CYP: *aða, aði, alli | aðak~ak, aðik~ayk, allik*³¹

– Dialects which generalize *ðu:* in the plural

- * NACHR: Pronominal forms: *ha:ða, ha:ði, ha:ðu | ha:ða:k, ha:ði:k, ha:ðu:k*; Adnominals: *ha:ð- | ða:k-*

²⁶ Here the Stage 2 process was probably further augmented by a re-segmentation of *haw-ki* and the analogous creation of *haw*, replacing original *hawl*.

²⁷ Here the proximal *hawda* has been extended into the plural, but note the biform *hawk*.

²⁸ Note the distal plural is similar to IQAQ though the source is a dialect map so we do not know if there are adnominal forms, or archaic forms as in that other dialect.

²⁹ Note the (probably later) gender marked distal masculine and feminine. The distal plural is somewhat unexpected.

³⁰ Note the borrowed proximal plural form, which is characteristic of many dialects near major population centers in the Levant.

³¹ Borg (1985, p. 142) reconstructs the plural forms to **hawla*. This dialect obviously does not fit the normal pattern of development of these dialects, but this will be discussed later.

- * NADJ: *ha:da, ha:di, ha:du* | *ha:da:k, ha:di:k, ha:du:k* | Adnominals:
ha:d- | *da:k, di:k, du:k*³²
- * NAFEZ: Pronominal forms *ha:da ha:di ha:du* | *ha:da:k (ha:)di:k*
(ha:)du:k; Adnominal forms: *ha:d-* | *da:k, di:k, du:k*
- * NAHAS: Pronominal forms: *(ha:)ðæ, (ha:)ði, (ha:)ðu* | *ða:k, ði:k,*
ðu:k
- * NAMz: Pronominal forms: *ha:da, ha:di, ha:du* | *(ha:)ða:k (ha:)di:k*
(ha:)ðu:k~(ha:)du:m Adnominal forms: *ha:d-*,³³
- * NATL: Pronominal forms: *ha:da, ha:di, ha:du* | *ha:dak, ha:dik, ha:-*
duk; Adnominal forms: *(ha:)de-, (ha:)de-, (ha:)de~(ha)du- | (ha:)da:k-*
*, (ha:)da:k-, (ha:)da:k~(ha:)du:k*³⁴
- * NATRJ: Pronominal forms: *ada adi ad^uun(i)* | *(a)dak (a)dik (a)d^uuk*;
Adnominal forms: *ad-*
- * NATRM: Pronominal: *ha:da(:ya), ha:di(:ya), ha:du(:ma(:ya))*³⁵ | *ha:-*
da:ka(:y(a)), ha:di:ka(:y(a)), ha:du:k-a/a:y/a:ya/umma; Adnominal:
ha: l-

³² This dialect also has a series *hauda, haida~haidi, hamda* (P. Marçais, 1956, pp. 461-462), with a corresponding distal sequence with *-k*. This is clearly derived from presentative+pronoun structures that produced *hahu~hau, hahi, hahum~hahəm~ham* which were analogously extended into the demonstratives by equating the initial *ha:-* prefix in the demonstratives with the presentative+pronoun structures (*ibid.*, p. 445). This is clear in the lack of agreement in the final *da*, where everything seems to be based on the masculine form. This is clearly a very different and hence independent process from that which produced the *ha:k* dialects, where we find no /m/ in the plural forms, for instance, or *hau* in the masculine.

³³ Jacques Grand'Henry (1976, p. 67) doesn't tell whether the shorter forms *ða:k*, etc. are actually adnominal, or not. It seems likely given the dialects in the area. Note that he believes the *(ha:)du:m* represents an older substrate.

³⁴ W. Marçais (1902, p. 118) claims that there is also a prefix *ħal-* but as W. Fischer (1959, pp. 45-6) argues, this is actually just a reflex of *had-*.

³⁵ There is also a series of proximal demonstratives which can be used as both adnominal and pronominal demonstratives. These are clearly derived from presentative **ha:+*pronouns: *a:hw^uá, a:hyá, a:h(u)m^uá*.

* NATUN: Pronominal *(h)aḏa(ya)*, *(h)aḏi(ya)*, *haḏum(a)~haḏula* | *(h)aḏaka~haka*,
(h)adika, *hadukum~hadumka*; Adnominal: *ha:* ~ *haḏ*

– Dialects which extended singular to plural with pronominal endings

* Y145: *ha:ḏa*, *ha:ḏi*, *ha:ḏum*, *ha:ḏe:n* | *ha:ḏa:k*, *ha:ḏi:k*, *ha:ḏunk*, *ha:ḏink*³⁶

* Y156: *ḏe:*, *ḏi:*, *ḏinne:n* | *ḏa:k*, *ḏe:k*, *ḏinne:k*

– Dialects which fused **ḏa:* with the plural demonstratives *?ula* and *?ila*

* Dialects which have **ḏa-?ila* > *ḏalla*

· ARQAUZ *ḏahah*, *ḏi:hah*, *ḏallahah* | *ḏa:kak*, *ḏikah*, *ḏalla:kah*³⁷

* Dialects which have **ḏa-?ila* > *ḏayla*

· AREM: *ha:ḏa*, *ha:ḏi*, *(ha)ḏe:l(a)* | *(ha)ḏa:k*, *(ha)ḏi:tf*, *(ha)ḏo:la:k*,
(ha)ḏe:la:k~(ha)ḏila:k

· ARANZ: *ha:y~(ha:ḏi)*, *(ha:ḏi)*, *(ha)ḏe:le~haḏo:l(e)* | *(ha)ḏa:k*, *(ha)ḏi:tf*,
(ha)ḏela:k

· ARBAH: *hay*, *ha:(de)*, *ha:(di)*, *(ha)dele:n* | *(ha(:))da:k*, *(ha(:))di:k*,
(ha(:))dela:k~hado:la:k; Also in a women's dialect, distal: *ha:k(a)*
he:k(a/i); Adnominal: *ha-*

· AROM: *(ha:)ḏa*, *(ha:)ḏi*, *(ha:)ḏe:la* | *(ha:)ḏa:k*, *(ha:)dik*, *(ha:)ḏe:la:k*³⁸

· YHADR: *ḏa*, *ḏi*, *ḏe:l(a)~ḏo:l(a)* | *ḏa:k*, *ḏi:k*, *ḏe:la:k*³⁹

* Dialects which have **ḏa-?ula*

³⁶ This is characteristic of a larger number of southern Yemeni dialects, but is the only dialect in the sample with this type of marking.

³⁷ There are other dialects of this type in Yemen.

³⁸ The **ha:* part is typically omitted when demonstratives are used adnominally.

³⁹ The confusion about the proximal plural demonstrative form is probably due to the fact that this is a huge area conflated into a single sample point.

- CHD: *da, di, dol* | *dak, dik, dolak*⁴⁰
- EGCAI: *da, di, do:l* | *dukha, dikha, dukham~dukhum~dukhumba*
- **Dialects which interpreted **ḏa-ʔila* as feminine plural forms**
 - i. CHDSH: *da:, di:, do:l, de:l* | *da:k, di:k, do:la(:)k, de:la(:)k*
 - ii. SHUK: Pronominal: *da, di, do:l, de:l* | *da:k, di:k, do:lak, de:lak*;
Adnominal: *ha-*
 - iii. NIG: *'da, 'di, 'do:l(a), 'de:l(a)* | *'dʕa:ka, 'dʕi:ka, 'dʕo:lak(a), 'dʕo:lak(a)*
 - iv. EGBR: *da~dih, di~diy, do:l(a), de:l(a)* | *dukka:ti, dikki:ti, dukkumʕmʕa, dikkinna*⁴¹
- **Dialects which added **ha:* to *ḏawla***
 - a. ARNJD: *(ha:)ḏa, (ha:)ḏi, (ha:)ḏo:l(a), haḏo:li* | *(ha)ḏa:k, (ha:)di:tf, (ha)ḏo:la:k, (ha:)ḏo:litf*;⁴² Adnominal: *ha-*
 - b. IQBGJ: *ha:ḏa, ha:yi, haḏo:li* | *haḏa:k, haḏi:k, haḏo:lak*; Adnominal: *ha-*
 - c. YWAS: *ha:ḏa, ha:ḏi, haḏowla* | *ha:ḏa:k, ha:ḏi:k*
 - d. LVDAM: *ha:da, ha:di~hayy, hado:l* | *hada:k, hadi:k, hado:li:k~hadənk*;
Adnominal: *ha-*
 - e. LVHEB: *ha:da, ha:di, hado:l(a)* | *hada:k, hadi:k(e), hado:la:k(e)*;
Adnominal: *ha-*
 - f. LV417: *ha:ḏa, hay, haḏo:l* | *haḏa:k, haḏi:k*
 - g. LV433: *ha:ḏʕa, ha:y, haḏo:l(a)* | *haḏa:k, haḏi:tf, haḏo:la:k*

⁴⁰ This dialect could have lost plural feminine markers, cf. CHDSH.

⁴¹ Distals clearly replaced by presentatives in this dialect.

⁴² For some reason, the *ha-* in the proximal feminine plural isn't marked as obligatory in Bruce Ingham (*EALL* : "*Najdi Arabic*"). It is unclear if this is a mistake or actually reflects the dialect they are describing.

- h. LVSPAL: *he:ðya, he:ðya, haðo:l* | (*ha*)ðe:k, (*ha*)ði:k, (*ha*)ðo:le:k⁴³
- i. LVSSOU: *ha:ði, ha:ði, (ha)ðol* | *haða:tf, haði:tf, (ha)ðo:latf*
- j. IQMosJ: *ha:ða, ha:yi, haðo:li* | *haða:k(a), haði:k(a), haðo:k(a)*⁴⁴
- k. LvS330: *ha:ða, ha:ði, haðo:l* | *haðe:k, haði:ke, hado:k*
- l. NABEN: *'haða~'hað'a, haði, ha'ð'o:l, ha'ðe:n*⁴⁵ | *ha'ð'ak(ki), ha'ðik(ki)*⁴⁶;
Adnominal: *ha-*
- m. SINR: *ha:ða, ha:ðiy~he:ðiy, ha:ð'al'(l'ah)~ho:ð'al'(ah)*⁴⁷ | *ha:ða:k, he:ði:k, hað'o:l'l'a:k*
- n. SINS: *ha:ða, ha:ðiy, ha:ð'al'(l'ah)~ho:ð'al'(ah)* | *ha:ð'a:k ha:ði:kih ha:ð'al'l'a:k~ho:ð'al'l'a:k*
- o. **Dialects which have **ha:-* attached to feminine plural **ða-* *ʔila* forms**⁴⁸
- i. IQBGG: *ha:ða, ha:ði~ha:y, haðo:la, (ha)ðanni* | (*ha*)ða:k(a), (*ha*)ði:tf, (*ha*)ðo:la:k⁴⁹
- ii. IQKHZ: *ha:ð(a), ha:ði~ha:y, (ha)ðo:l(a), (ha)ðanni* | *ða:k(a), (ha:)ði:tf, ðo:la:(a), (ha:)ðannitf*
- iii. LvS281: *ha:ða, ha:ði, haðo:l, haðanna* | *haða:k, haði:tf, haðo:la:k, haðannitf*

⁴³ These forms are from Behnstedt (1997). Cantineau (1934, pp. 219-220) gives *he:ðe, he:dei, daðo:l* | (*ha:*)ðek, (*ha:*)ði:k, (*ha:*)ðu:lek, where the feminine proximal has a very slight off-glide.

⁴⁴ Note the distal plural form. Might imply that the demonstratives in this dialect are borrowed later.

⁴⁵ The providence of the feminine plural is unclear.

⁴⁶ Owens (1984, p. 54) claims there is also a dual form *haðik'ke:n* that is used rarely.

⁴⁷ The changes in the initial syllables here appear to be phonologically motivated.

⁴⁸ This assumes that *ðanna*-type forms are originally **ðalla*.

⁴⁹ The feminine plural form is marginal in this dialect.

iv. LvS340: *he:ða, ha:y, haðo:la, haðanna | haðe:k, haði:ki, hadawle:k*

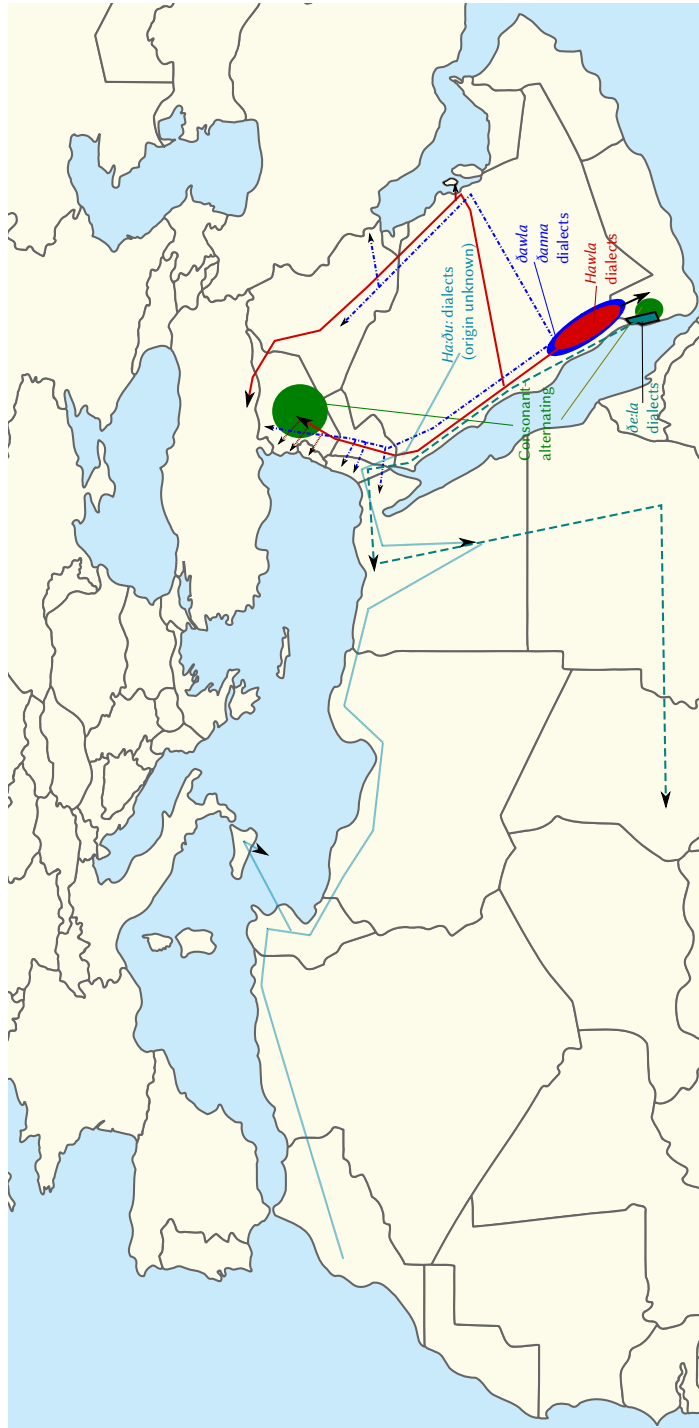
- Dialects which are difficult to classify
 - AFGA: *had, hadi~hay*⁵⁰
 - MALT: *da~dan(a), di~din(a), daw~dawn(a) | dak(a), dik(a), dawk(a)*
 - UZBA: *ha:d~hat ~ha:z, ha:di:, hadla:u, hadla:n | du:k~duk, duki:, dukala:u, dukala:n; Adnominal: ha-*
 - UZBDJ: *ha:d~hat ~ha:z, hai, halo: ~hala:u, hala:n | du:k~duk, diki:, dukala:, dikala:n~dikila:n*

Figure 6.4 is an attempt to summarize this information, while showing the extent of the alignment events that created the pre-Islamic dialect groupings. The only development in Arabic which really seems to have operated in a tree-like manner, with a clear division between two groups, is the split between vowel and consonant-alternating dialects, and this is therefore represented by the split at the top of the diagram, and the division of the page between the vowel-alternating dialects on the left, and the consonant-alternating dialects on the right, with a dashed line dividing the two groups. There are of course alignment events which later affected dialects from both groups, and so these cross that line. Boxes with solid lines show the extent alignment events which clearly seem to have happened in a single place. Boxes with dashed lines group dialects which have superficial similarities, but which are not necessarily the result of a single alignment event. There is no difference between the amount of roundness in the boxes.

⁵⁰ Data for this dialect's demonstrative system is very limited.

Dialects which arose through a process of koineization or dialect mixing are placed in their current dialect group, but lines with circles at the end show the input dialects. Dialects which show different layers of demonstratives are shown in their current group, but have a line with a square on the end showing their original dialect type.

Finally, Figure 6.5 is an initial attempt to place the speech communities described in this chapter onto a map. This is primarily as an aid to the reader, and should not be taken to be a definitive portrayal of the early Islamic speech communities. Solid lines represent pre-Islamic movements, while dashed lines represent post-Islamic movements. An attempt has been made to use different kinds of dashed lines to distinguish the different communities and their movements.



Conclusion

This dissertation has shown that a speech communities framework for historical reconstruction can reveal a great deal of information about the history of the Arabic language, in spite of the challenges of high dialect diversity and a long history of contact between those dialects. The dissertation proposed that framework based on research in historical linguistics and sociolinguistics, emphasizing the role of speaker attitudes and allegiances as a determiner of the spread of linguistic features. Since this approach requires a firm grounding in the historical circumstances of the development of Arabic, the dissertation analyzed the human geography and history of the Arabic speaking world. It then analyzed the linguistic history of Arabic demonstrative pronouns and adnominals. Finally, the dissertation integrated the historical and linguistic information within the speech communities framework to show what Arabic-speaking communities existed prior to the Islamic conquests and how they are related to modern Arabic-speaking communities. This conclusion will go into further depth on several questions that have been raised by the dissertation or are currently being debated in the field of Arabic linguistics.

The following is a brief summary of the major findings of the dissertation, organized in order of relative importance and centrality to the goal of the dissertation. Issues discussed later in the conclusion are not included in this summary.

- Early Arabic demonstratives marked no gender in the singular (only *ḏa:* for masculine and feminine), and so Arabic dialects innovated two primary methods of marking gender, based on analogy with pronominal and verbal forms: Consonant-alternation with m., f. **ḏa:*, **ta:* and vowel-alternation with **ḏa:*, **ḏi:*. (Section 5.3.1.1)
- Early Arabic plural demonstratives varied in the vowel of the first syllable, between *ʔula* and *ʔila*. This variation should be reconstructed to Semitic. Dialects with vowel-alternation in the singular demonstratives largely have plural *ʔula*; consonant-alternating dialects largely have *ʔila*. (Section 5.3.2.1)
- The **ha:-*prefix of the demonstratives was originally a presentative, and unlike other Semitic languages was not affixed to the demonstrative during the process that produced definite articles. (Section 5.4)
- Demonstratives of the type **ḏa:-ʔila* and **ḏa:-ʔula* were calqued based on the demonstratives with the **ha:-*prefix of the form *ha:-ʔula*. (Section 5.3.2.2)
- The dissertation identified the following major pre-Islamic Arabic speech communities:
 - Consonant-alternating dialects. It is not clear where they arose, but they were prominent in the fourth century Levant and are found today in the Yemeni Tihāma. (Section 6.1)
 - *hawla*-dialects, with plural forms derive from *ha:-ʔula*. These probably arose in the southern Hijaz early in the common era, but replaced the consonant-alternating dialects in the Levant in the fifth or sixth century. They are widespread, and represent the older stratum of dialects in Iraq and the Levant and may be attested in Andalusian and

Bahrani dialects. They are well established in the northern Yemeni plateau. (Section 6.2)

- Dialects which have reallocated **ḏawla* and **ḏayla* to mark masculine and feminine plurals respectively. These dialects probably arose in north-western Yemen. They are primarily attested in the dialects of the Nile Valley. (Section 6.3.2)
 - Dialects with masculine plural **ḏawla* and feminine plural **ḏanna* which also arose in the southwest Arabian Peninsula. These dialects were probably spoken in the *'amṣār* of Iraq, and represent a relatively later stratum of dialects in Iraq and the Levant, particularly nomadic dialects. (Section 6.3.3)
 - Dialects with plural *ha:ḏu*, whose origins are unclear. These dialects first moved into Upper Egypt, and then into North Africa probably in the tenth or eleventh centuries. By the fourteenth century, most of these dialects in Egypt had shifted towards the dialect of Cairo. (Section 6.4)
- Classical Arabic forms appear to reflect the different speech communities that were prominent in the Hijaz and the Levant over time. (Section 6.6)
 - The standard narrative of the Hilalian invasions, and their influence on the dialects of North Africa seems to be inaccurate and should be revised. (Sections 4.2.1 and 6.4)
 - The apparent conservatism of nomadic Arabic dialects is actually a result of the mobility of nomads and their insulation from linguistic changes beginning at the top of the urban hierarchy. (Section 2.5)

6.8 Evaluating the Speech Communities Approach

This section seeks to evaluate how successful the speech communities approach has been, as applied in this dissertation.

It is difficult to thoroughly evaluate the approach when only one linguistic variable, demonstratives, was used in the reconstruction. Using only one variable reduces the *prima facie* validity of the study, and a single variable eliminates the problem of contradictory, cross-cutting isoglosses. Nonetheless, the linguistic reconstruction has, at times, been corroborated by historical data. Moreover, I predict that the speech communities reconstructed here would have innovated other forms outside of the demonstrative system, and that those forms would have diffused within the limits of those same communities. I will be able to test this prediction with later research.

We can evaluate the speech communities approach by comparing it with the more established comparative method as it is traditionally applied. There are two primary areas where the speech communities approach differs from the comparative method: First, the speech communities approach focuses not on the ‘ultimate origins’ of features, but rather what might be called their ‘proximate origins’, the most recent and relevant speech community that was characterized by those features. Second, the speech communities approach focuses on the role of speaker attitudes in language change in order to show how speech communities interacted with one another.

Certainly traditional reconstruction and subgrouping would have produced a categorization of the modern dialects quite similar to one produced here. An approach that was concerned only with the ultimate origins of the demonstratives would likely find only disappointing results, however. Like most Semitic languages, Arabic probably inherited demonstratives of the type δi -, $\delta a:(t)$, ʔVla ,

but the ultimate ancestor of every modern dialect (i.e. ‘Proto-Arabic’ probably had generalized the singular to a single demonstrative $\delta a:(t)$ which tells us very little about the later diversity.

Indeed, it is not clear exactly what is gained by referring to Proto-Arabic. If one hopes to do deeper reconstruction, it might be a ‘convenient fiction’ in the words of Henning Andersen. But it would be untenable to equate Proto-Arabic (or Proto-Araboid) with the variety that first ‘split off’ from Central Semitic, since there was clearly continued influence between early Arabic and the other Central Semitic varieties. For example, the fact that the **ha:-*formative became a demonstrative prefix in many varieties, instead of an independent demonstrative, may be due to influence from languages where the phonetically similar (and possibly etymologically related) definite article was affixed to the demonstratives. The apparent similarities between the two forms may have tipped the scales towards the formation of **ha:-*prefixes rather than independent demonstratives, though both were equally likely developments.

Another important outcome of this study is the distinction between time and place of *origin*, and time and place of *sociolinguistic prominence*. We have no reason to believe that consonant-alternating dialects developed at a different time than vowel-alternating dialects, and certainly neither need be derived from the other. However, our data very clearly shows that consonant alternating dialects were prominent and important much earlier than vowel-alternating dialects. There is nothing in the linguistic structure of these dialects which explains their successes at particular times; instead, they were carried by their speech communities and the fate of these demonstrative forms is related to the successes and failures of those who spoke them.

This approach also gives us the tools to understand the complexities of

layering and contact that gave rise to these dialects. A dialect might have proximal demonstratives of the vowel-alternating type, and distal demonstratives of the consonant-alternating type. This implies that that dialect ultimately originated from a consonant-alternating dialect, but clearly it has undergone significant influence from a vowel-alternating dialect. By searching for more than just the ultimate origin of the dialect, we can begin teasing apart the historical and linguistic forces that lead to the development of that and other dialects, and so build a more complex understanding of the history of Arabic.

Finally, by clearly situating the reconstructed speech communities in time, it is easier to understand how they were related to one another, and how they relate to modern dialects. This contrasts with the approaches of both Al-Jallad (2012) and Cowan (1960). It is not clear whether exactly Cowan's 'Proto-Western Arabic' corresponds to any of the speech communities reconstructed here, and whether it includes features that were layered due to the successive migrations into North Africa. Al-Jallad's 'Ancient Levantine Arabic', as we have discussed, clearly includes different layers of development — for example, he reconstructs a relative marker *ta*: which is really just a remnant of consonant-alternating demonstratives, but reconstructs vowel-alternating demonstratives — and it is not clear exactly how to compare that reconstructed variety with other reconstructions or with the speech communities reconstructed here.

The focus on the attitudinal factors in the diffusion of linguistic features proved to be a very successful part of the speech communities approach. We are able to discern linguistic attitudes in great detail due to an unusual aspect of contact between closely related languages. In contrast to more distantly related languages, where borrowing is easy to discern, in closely related languages 'borrowing' a form is essentially identical to diffusion, which is diagnostic of participation

of the donor and recipient dialects in a single speech community. In contrast, closely related languages can also choose to calque a form from another dialect, showing that they participate in the same speech community as the donor dialect, but retain their independence or otherwise find the donor dialect less prestigious than a dialect which borrows the form wholesale. In the words of Kahane (1986, p. 503), calques are a “the product of linguistic loyalty” to one’s closest speech community which “purify the clothing but not the body”, a very different statement of linguistic belonging than accepting foreign forms via diffusion.

By acknowledging the role of attitude in linguistic diffusion, we can therefore diagnose to a very fine degree the relationships between different speech communities. Furthermore, contact between closely related dialects will always be an integral part of the process of language differentiation, so these tools for gauging the relations between speech communities are widely applicable in historical linguistics.

6.8.1 Tribes as linguistic units

In Section 2.4, I argued that the tribe is not *necessarily* a linguistic unit, though it theoretically can be if it fulfills the criteria for a speech community of acting as a focus point for shared community allegiances, and most importantly, as being sufficiently connected by network ties. Many of the tribal units reported in the indigenous Arabic historical tradition do not seem to fulfill both of these criteria. The tribe of Tamīm, for example, apparently had branches throughout the Greater Arabian Peninsula, which means they probably lacked sufficient network ties for changes to diffuse throughout all those who fell under the Tamīm umbrella. Tamīm may not even have been a contemporaneous political unit. In the grammatical literature, for example, Tamīm is contrasted to Hijāzi, an opposi-

tion of a clearly geographical unit to an apparently tribal one (Rabin, 1951, pp. 1-3). We certainly need to investigate the complex linguistic ideologies which gave rise to this opposition, but in the meantime we cannot be sure whether Tamīm really represented an emic ideological center of allegiance.

On the other hand, we have one very clear instance in the data where a tribal group corresponds to a speech community. This is the case of the tribe of ‘Akk, which is said to have been the primary colonizing group in Egypt (see further Section 6.3.2). In this case, we have a relatively small tribe (said to number around four thousand) whose geographical extent was quite limited (a small stretch of the Yemeni Tihāma). We have very clear information their role in the Islamic conquests, which state that this specific tribe was a major founding group of Fustāt. Our linguistic information on this tribe is based on reconstruction and on the modern dialect of the region, not on reports in the grammatical tradition, and so is independent of the historical-grammatical tradition.

Here then is a successful uniting of linguistic and social information, bound by the social unit of the tribe. This suggests that the tribe can be a linguistic unit if we can be relatively sure that it is small enough to be united by network connections and the definition of that tribe is probably emic and not imposed from outside. This is also an excellent example of how the speech communities approach can link linguistic and non-linguistic data.

6.8.2 Semitics and Arabic

This dissertation provides further evidence that reconstructions of Semitic must consider the diversity of modern Arabic dialects, as well as the diversity of past language states. Hasselbach’s (2007) reconstruction of the plural demonstratives in Semitic as having an original /u/ clearly does not match the Arabic

distribution, and there is no reason to assume homogeneity in early Semitic. Sampling widely among Arabic dialects allows for a better understanding both of the Araboid branch of Semitic, and probably Semitic more generally.

This dissertation has also shown that Classical Arabic must be treated with a great deal of caution in historical reconstruction. Even if it drew directly on a single spoken dialect, which is doubtful, that dialect may have been composed of different, complex layers that date to different eras, reflecting different speech communities.

We must also remember that the codifiers of Classical Arabic were as much linguists as we are today, and did their best to make sense of chaotic data. They were developing a particular system of analysis, and forms were, at times, made to conform with their analysis. The best example is the situation analyzed in footnote 34 on page 300, where grammarians had reports of a relative form *?allaðu:na* and assumed that this must be a nominative form, with an oblique, *?allaði:na*. However, the original report of this form, in al-Farrā' ([d. 822] 1983), was actually using *?allaðu:na* to illustrate case-invariable dual and plural forms.

Linguists of that time, and of this (Wright, 1896-1898, § 347, Rem. b; also in Al-Jallad, 2012) have been all too happy to accept the explanation of the *?allaðu:na* as an example of a 'nominative' relative, since that conforms to their understanding of the early Arabic case system. However, the research here would suggest that such forms may be related to the North African and Yemeni demonstratives of the form **ha:ðun* whose final syllable is derived from analogy with the pronouns, not the plural endings of Classical Arabic.

Finally, the results of this dissertation suggest caution when applying Hetzron's (1976) principle that "borrowing in the morpholexicon [...] happens in very special circumstances only, and rarely enough that it would not practically weaken

the strength of shared morpholexical innovations as a criteria for subgrouping.” However, between closely related languages diffusion and borrowing are almost identical, even for morpholexical forms. Since every language undergoing differentiation has long periods of contact between closely related languages, this distinction is only really valid at a quite late stage in a language splitting scenario. Indeed, even if we only consider the first diffusion event in which a form spread, and treat later events separately, demonstratives are borrowed quite often between different Arabic dialects and so are not a very strong criteria for subgrouping in the classical sense of the comparative method.

Furthermore, the real determinant of whether a form is or is not borrowed (or diffused) frequently or easily is its sociolinguistic importance. In Arabic, demonstratives seem to be a major site of linguistic construction of community — apparently more so than in other Semitic languages, which have much less diversity in their demonstratives than Arabic. The diversity of distal demonstratives in Egypt illustrates this point, with villages separated by quite short distances having wildly different demonstratives. Borrowing, not borrowing, or calqueing a demonstrative from another Arabic dialect is therefore a powerful expression of linguistic unity or disunity, in a way that other linguistic features may not be.

6.9 Out of Yemen or Out of Levant

One of the major arguments advanced by Al-Jallad (2012) is that modern colloquial dialects hail not from Yemen, as is claimed in the Arabic historical tradition, but rather directly out of the Levant. That is to say he argues that Arabic dialects ultimately derive from ancestors that were spoken in the Levant, and not

in Yemen.⁵¹

Certainly Arabs are attested in the Levant (and Egypt, Iran, and elsewhere within the Middle East) as early as 853 BCE, when Arabs are recorded participating in a battle in north-western Syria, near the modern border with Turkey (Macdonald, 2009a).⁵² On the other hand, Araboid languages (referring to both ANA and Arabic, as described in Section 3.3.1) are attested throughout the strip between the Levant and Yemen. ANA inscriptions in these area, particularly near the Levant, are attested from the eighth century BCE to the fourth century CE. Arabic inscriptions begin in the fifth century, though there is certainly reason to believe that Arabic as such was spoken for much longer (Macdonald, 2010). Even Al-Jallad (2012, p. 379) admits that there could have been movement southward from the Levant into the Peninsula by Arabic speakers within the first few centuries CE.

The question then is the same one of ultimate versus proximate origins. Claims about an ancient original *urheimat* are not necessarily any more helpful than claims regarding an original proto-language, since we are interested in more recent history (see Section 1.1.2).

The truly ancient history of Arabic in the Levant would only be relevant if, as Al-Jallad (*ibid.*, p. 379) claims, the modern Levantine dialects really did “continue the ancient varieties.” However, both our linguistic and non-linguistic evidence cast doubt on how ancient any modern attested Levantine dialect could be. As argued in Chapter 2, areas with low population density are prone to popula-

⁵¹ Another recent article, Kerr (2013), makes a similar assertion about the origins of Quranic Arabic and so one wonders whether this is becoming something of a fashionable argument to make.

⁵² The text mentions an Arab ruler, Gindibu, who brought one thousand camels to the aid of the Assyrians, suggesting that though the battle was in fertile areas near the Mediterranean, Gindibu or those he had control over, were probably part of a camel-rearing group living in steppe-lands, since camel rearing is of limited utility within sedentary well-watered areas.

tion replacement, and thus an apparent ‘flip’ in dialects. In Section 3.4, I show that Arabs tended to live in the areas of the Levant which were necessarily of lowest population density. This changed around the sixth century CE, at which time they may have begun to expand at the expense of Aramaic speakers who had inhabited more fertile areas. The groups that would have undertaken the Arabization of the Levant would probably have been nomadic or semi-nomadic, and so hailed from areas where apparent population replacement would happen quite easily. Therefore, the proximate origins of coastal and mountain Levantine dialects is really the dialect milieu of the Syrian desert of the sixth century, and probably did not reflect some ancient stratum.

Indeed, we have good evidence from the demonstrative systems in the Levant, chronologized by the Namāra inscription, that there was a significant change in the demonstratives used in what are now rural Levantine dialects (which Al-Jallad focuses on), with the older demonstrative forms only preserved in marginal structures (see Section 6.1). It would be remarkable for such a change not to be accompanied by a variety of other linguistic changes, and so many other ‘ancient’ features might also be lost or attested only in the most marginal contexts. The city dialects of the Levant show yet another (and I argue, later) layer of dialects, so even if, for example, the Psalm Fragment is, as Al-Jallad (2012, pp. 144-9) argues, originally a fourth century document, it does not necessarily record a direct predecessor to the (urban) dialects that followed it.

There are other linguistic features which also suggest a more complex picture than a ‘Levantine dialects in-situ’ scenario. All Levantine dialects now show an definite article *il-* where the *l* assimilates to following coronals, as is true throughout Arabic dialects with the exception of some Yemeni dialects. However, according to the analysis by Al-Jallad (*ibid.*), the definite article did not show assimilation

in old attestations of Arabic in or near the Levant. The clearest example of an assimilating article is in what he categorizes as an ANA text, the Rbbl inscription found at Qaryat al-Faw, and dated to perhaps the first century BCE. If we follow his essentialist approach to categorizing this inscription, we clearly have a form of the definite article that is present in the south-western peninsula at this early date, and which was sociolinguistically important enough at that time to penetrate an entirely different language.⁵³ Somehow, in the modern era, all of the dialects of Arabic outside of the south-west Arabian peninsula have an assimilating definite article, which suggests that some group must have deeply influenced the Levantine dialects at a later time than is reflected in non-assimilating inscriptions.

Indeed, given Al-Jallad's (2012) emphasis on textual attestations of Arabic, and the importance of these attestations (especially the Psalm fragment) to his reconstruction, it seems unlikely that the time depth of his Ancient Levantine Arabic is much deeper than that of his oldest attestation (the Namāra inscription from 328 CE).⁵⁴ Hence, even the Arabic attested in the Levant might not be any older than the first few centuries CE, and could certainly have, at some point, come from further south.

There are two other arguments Al-Jallad (2012) makes against the 'out of Yemen' theory. The first is that the accounts within the historical and genealogical tradition of Arabic emphasize the Yemeni (or following the convention from Chapter 3, Yamani) origins of various groups, and claim that there were various migrations out of Yaman northward. Al-Jallad (*ibid.*, p. 24) that claims Yamani

⁵³ Obviously the situation is more nuanced than this, and the distance between the ANA languages and Arabic would be relatively small, so the diffusion of this feature would not necessarily have been difficult. Nonetheless, it shows that there was a shared speech community in this area which had this feature.

⁵⁴ Though to be sure, he also references some onomastic data that is older than this, but neither the data nor the arguments are abundant enough to be decisive.

links were probably “the work of Yemeni patriots who were seeking a position of influence within a society that so valued genealogy.” First, it is not clear why they would choose to link themselves to a Yamani homeland. They could just have easily laid claim to a relationship with the various dynasties which had been dominant clients of the Romans or the Persian empires. A complex historiographic argument is needed to dismiss these claims as “mythological elements” that cannot be treated as a historical narrative. Such an argument would be of great interest, but it is certainly outside of the scope of either this, or Al-Jallad’s, dissertation.

Second, it’s important to understand that in early Islamic sources, Yaman simply referred to much of western Arabia south of Mecca, often meaning those areas outside of the control of the Byzantines (Bashear, 1989), so a Yamani affiliation need not mean that a group hailed from the modern state of Yemen, but rather from a significant portion of the Arabian peninsula, and certainly an area that was among the heartlands of early Islam.

Moreover, as argued in Chapter 2, Yemen was probably one of the most populous areas of the Arabian peninsula, and most likely to be able to provide surplus bodies for the Islamic military campaigns. Fertile areas of the Levant or Iraq may also have provided soldiers, but these were largely under the control of groups who opposed the early Islamic conquests.⁵⁵

Finally, we also have clear evidence of Islamic era migrations where the speakers are claimed to come from Yaman and this claim is corroborated by our linguistic evidence. The historical sources claim that the tribe of ‘Akk, hailing from the Yemeni Tihāma, were the major founders of Fustāt and hence of the modern

⁵⁵ It is not clear why Al-Jallad does not propose an ‘out-of-Iraq’ theory, since we know that Arabs had penetrated deeply into northern Iraq at various times in history, and the *qeltu* dialects which he claims are descended from Ancient Levantine Arabic are more widespread in what is now Iraq.

Egyptian dialect. Both modern synchronic dialect data and our reconstruction support the historical accounts, as detailed in Section 6.3.2, and so we do have a clear example of a tribe with claimed Yemeni roots whose dialect played an extremely important role in the Islamic conquests.

Thus, regardless of the antiquity of Levantine dialects, Yemeni dialects played an important role in the Islamic conquests. This returns to the question of chronology, since the six hundred years or so that al-Jallad allows for the presence of Arabic dialects in the southern peninsula (and I argue for a somewhat longer period of time in Section 3.3.3) were more than enough for significant linguistic changes to occur among those dialects, which would then be brought to other areas by the Islamic conquests, as happened in this case. Thus, we have a question not of an *urheimat* but of multiple places of proximate origin.

The final argument used by Al-Jallad (2012) against the ‘out of Yemen’ hypothesis relies on al-Hamdānī’s account of the linguistic situation in Yemen in the tenth century CE. In this account, al-Hamdānī describes where Himyaritic, as opposed to Arabic, is spoken. Even if we at face value take his account of, for example, “bad Arabic”,⁵⁶ the actual area that is not Arabic speaking in Yemen in the 10th century is quite small, and restricted largely to the highest elevation areas, as shown in Figure 3.3 on page 165. Himyaritic, whatever its linguistic status exactly, is certainly not, as Al-Jallad (2012, p. 198) would have it, the “predominant” language of Yemen in the tenth century.

Furthermore, it seems implausible that Yemen would be so heavily Arabized by the tenth century if that process only began in the sixth century. Yemen contains the highest and most difficult terrain in the Arabian peninsula and it has

⁵⁶ This appellation probably reflects deeper and more complex concerns about performative register and not just mutual incomprehensibility (see also C. J. Robin, *EALL* : “*Himyaritic*”).

been a political backwater throughout the post-Islamic era. In comparison, the Levant, which has relatively lower-elevation terrain,⁵⁷ and which was at the heart of the Arabo-Islamic world for much of its history, still has pockets of Aramaic speakers (not in the mountains, but in the rolling steppe outside of Damascus). Given the greater barriers to Arabization in Yemen, we expect a relatively earlier commencement date for the Arabization process in Yemen than in the Levant, and certainly it must have begun quite a bit earlier than the sixth century.

The verdict, then, on the ‘Out of Yemen’ or ‘Out of Levant’ accounts for the origins of Arabic is that they are essentially a false dichotomy. Population movements certainly would have gone in both directions, and the question of the ‘origin’ of Arabic is only really meaningful with clearly defined chronological parameters. In the study here, we are not concerned with the ultimate origin of Arabic, if that is indeed discernible or useful, but rather we are seeking to determine what kind of diversity existed among Arabic dialects in the pre-Islamic Arabic-speaking world, and if possible, where that diversity was located. As we have seen here, the answer to that question may indeed be the Levant, for *hawla-ha:k* dialects, though we cannot rule out a northern Mesopotamian origin, or the Levant-Iraq desert area. For other dialects, such as what are now the primary Egyptian dialects, the proximate place of origin is Yemen, and many other developments probably arose in the southern Arabian peninsula. For other dialects, such as the North African **ha:ðu:* dialects, the ultimate origin is unrecoverable, but we do know that these dialects spent some time in Egypt in the post-Islamic era, during which time they underwent further developments which are distinctive to those dialects.

⁵⁷ Lebanon’s highest peak is some 3,088 meters, as compared with Yemen’s 3,666 meter Jabal an Nabi Shu’ayb, though this belies the greater area and elevation of Yemen’s highlands versus the Lebanon and anti-Lebanon mountains.

6.10 Successes and Failures of Arabic

The model of language spread developed in this dissertation goes a long way towards explaining why Arabic was successful (or not) in many of the regions that it spread to as a result of pre-Islamic and Islamic population movements. The only other major work which has investigated this question is Ostler (2005), who attributes much of the success of Arabization to the fact that many of the languages spoken in areas conquered by Arabic were from the Afro-Asiatic language family. The results here suggest instead that it was a combination of population density, propitious timing, and possibly subsistence patterns.

The primary factor in the success of Arabic seems to have been population density, in both a local and general sense. The Arab conquerors were very careful about establishing new, separate settlements (*'amṣār*), exploiting what we now call the principle of first effective settlement. Even with years of plague and warfare, most of the countries that the Arabs moved into had relatively dense populations speaking other languages. The *'amṣār*, which quickly became the economic and political centers of the conquered regions, acted as linguistic beachheads, encouraging non-Arabs to move into the new, Arabic speaking cities. Migrants to those cities would be much more likely to assimilate linguistically than their native inhabitants, and certainly by the second generation there would be ample opportunities for childhood acquisition of Arabic.

In some ways, the physical choice of new urban environments was like the *mawlā* system, where in the early Islamic world new converts were integrated directly into the tribal system (Crone, EI2, "*Mawlā*"), taking on Arab names and a specific tribal identity. Just as these new members of the Islamic polity took on an Arab identity, new city-dwellers became part of a Arabic-speaking urban milieu and would be expected to accommodate to the norms of that new, Arabo-Islamic

urban society.

We have a clear example of exactly how strong the first effective settlement principle in the case of Egypt, where a small group of some 4,000 speakers set the linguistic behavior of the urban centers of Egypt for a millenium and a half, as detailed in Section 6.3.2. Indeed, the influence of Fustāṭ and then Cairo was strong enough that the urban demonstratives erased the demonstratives used in the countryside, which are now preserved only in papyri and marginal forms.

The vagaries of fate and timing also contributed to the success of the Arabic language. The bloody wars between the Byzantines and the Sassanians disrupted settlements from Iraq to Egypt and plagues devastated the same areas, immediately preceding and contemporaneous to the Islamic conquests. In the Levant this depopulation was key to the process of Arabization (see Section 3.4), but the disruption and depopulation would have smoothed the way for the Arabization elsewhere. Where there was little effect of either war or plague, as in the areas east of the Zagros mountains (i.e. most of modern Iran), Arabization was much less successful.

Nomads, with their high mobility, also must have played some role in the spread of Arabic, and at the same time might explain the lack of Arabic penetration into Iran. Nomadic Arabic speakers successfully colonized Khuzistan, the plains at the base of the Zagros mountains, but they do not appear to have moved further east into the higher altitude. Even if they did move into those areas, it would be difficult for the newcomers to compete with mountain-dwelling nomads who were already familiar with the appropriate seasonal migration patterns to different elevations (Donner, 1989). The same might be true of North Africa, where Berber speakers held out the longest in high altitude areas.

Given the established role of population density and subsistence pattern

in the success of language spread, it seems unnecessary to appeal to linguistic similarity or dissimilarity as Ostler (2005) does. Instead, we should look at places where Arabic was apparently a ‘failure’ to see what factors may have caused that failure.

In Andalusia, we do not really know how established Arabic was, though the evidence suggests that Syrian *hawla* dialects were successfully maintained by Arabic speakers, even if they did not necessarily Arabize non-Arabic speakers. Arabic may also have been poorly established in Andalusia as a result of the scattered and heterogeneous settlement patterns in Andalusia that did not necessarily create strong *ʿamṣār*. However, the primary reason Arabic isn’t spoken today in Andalusia is not the mismatch between Arabic and Romance, but rather the Reconquista which effectively removed most Arabic speakers from the Iberian Peninsula.

In contrast, in North Africa, Arabic has been most successful precisely in the places where Romance speakers probably resided, that is to say in the coasts littorals between the mountains and the sea. Though we know there were Romance speakers in these areas, there are none (from that stratum) found there today, while Tamazight speakers have much more successfully resisted assimilation to Arabic, in spite of their greater ‘genetic similarity’ to Arabic.

Iran also does not speak Arabic, except in the plain of Khuzistan which is west of the Zagros, though it was under Arabo-Islamic rule for centuries. This is likely due to the high population density of Iran at the time of the conquests, which had been spared the plagues that struck Iraq and Syria Conrad (1981, p. 338). The barrier of the Zagros mountains (and the fact they were already inhabited) would also block the movement of Arabic-speaking desert nomads. Persian cultural continuity must also have played a major role in the non-Arabization of Iran,

especially as Arab culture became more Persianized in Iraq.

Finally, it is important to dispense with the notion that all acquisition of Arabic by non-Arabs was somehow adult second language acquisition, and not native-like child or adolescent acquisition as is often argued by studies on the history of Arabic, including Ostler (2005). Huge numbers of slaves were raised in Arabic-speaking households, and they would have had ample the opportunity to learn Arabic as a native or near-native language. Similarly the *mawlā* system thoroughly integrated non-Arabic speakers into an Arabic-speaking community, so that within a generation they would likely have had the opportunity for childhood acquisition.

The *ʿamṣār* also provided ample opportunities for non-native speakers to acquire the dominant language of the *miṣr* and we even have literary reports of this. The poet Baššār b. Burd was born to a Iranian freedman who had been moved to Basra as a slave. He was a *mawlā* of the Arab tribe of ʿUqayl, and apparently acquired both spoken and literary Arabic as a native speaker, beginning his poetic career at age ten. One anecdote holds that Baššār was reciting poetry among a group of Arabs, who were unable to tell whether he was an Arab or a *mawlā*, rather convincing evidence that he spoke with a native-like accent (al-Iṣfahānī, 1969-1982, 3:1012-13).

The belief that Arabic was largely acquired by adult learners is therefore not particularly tenable. Indeed, a large source of the success of Arabic was the policy of founding *ʿamṣār* which placed non-native speakers in a milieu where they would be likely to acquire Arabic natively, while slavery and the *mawlā* system increased the critical mass of Arabic speakers, a type of local population density. Therefore, non-Arab populations probably had little difficulty acquiring Arabic, regardless of the linguistic distance between Arabic and their native language, and

this is borne out by the evidence we've presented.

6.11 Further Research

One of the first steps for further research would be to extend this analysis to other linguistic variables beyond the demonstratives, though we cannot expect that every variable diffused at the same time through the same speech communities. Nonetheless, there are already some positive signs that the speech communities reconstructed here do correlate with other classifications. The *hawla* dialects correspond fairly well with both the Iraqi *qeltu* dialects and the rural Levantine dialects, which Al-Jallad (2012) argues are both ultimately descended from 'Ancient Levantine Arabic.' Though we have discussed some issues with that reconstruction, it is encouraging to see that multiple innovations appear to have diffused through the same speech community that innovated *hawla* demonstrative forms.

This study has run into difficulty with the major lacunae in our knowledge of Arabic dialects. Though we have detailed maps of Syrian, Egyptian and Yemeni dialects, our knowledge of Peninsular dialects especially those of the central Hijaz, and of the eastern Peninsula is very poor. Similarly, Iraqi dialects are not well described, with robust dialect data available only for a handful of cities and villages in Iraq. More field research is needed in all of these areas before we can be confident in the results of our linguistic reconstruction of the history of Arabic dialects.

Many of the articles that we do have on these dialects are actually short sketches, with inconsistent coverage of forms. It might be helpful to create some type of database that collects whatever linguistic material is presented in a variety of different articles, similar to the dialect sketches in the Encyclopedia of Arabic Language and Linguistics, but in a searchable, geographically situated format.

Such a database would be helpful both for Arabists and Semiticists, since it would make dialect descriptions more accessible and easily searchable.

Finally, further research is needed on the process that created Classical Arabic and how it relates to the early spoken dialects. As we have increasingly detailed reconstructions of pre-Islamic Arabic, it becomes easier to see where Classical Arabic has directly borrowed certain forms, and where new forms were created that are specific to the literary register. Understanding that complex and dynamic process can help explain how literary registers evolve more generally, and to what extent they preserve old features, and to what extent they innovate entirely new ones. A changing understanding of Classical Arabic, along with a better understanding of the history of the dialects, may also force us to reconsider how we reconstruct Arabic and Semitic.

Appendix

Table A.1: Demonstrative data from sample: Proximal forms

Dialect name	M.SG	F.SG	M./C. PL	F. PL
AfgA	had	hadi hay		
And	'haða (ða~ði)	('ha)ði	'hawlin(k)~'hawl(ay)	
AnDar	a:za	ayzi	o:zi	
ArAnz	ha:y~(ha:)ði	(ha:)ði	(ha)ðe:le~haðo:l(e)	
ArBah	hay ha:(de)	ha:(di)	(ha)dele:n	
ArEm	ha:ða	ha:ði	(ha)ðe:l(a)	
ArNjd	(ha:)ða	(ha:)ði	(ha)ðo:l(a)	haðo:li
ArOm	(ha:)ða	(ha:)ði	(ha:)ðe:la	
ArQah	ðah	tah	?awði	
ArQz	ðahah	ði:hah	ðallahah	
Chd	da	di	dol	
ChdSh	da:	di:	do:l	de:l
Cyp	aða	aði	alli	
EgBr	da~dih	di~diy	do:l(a)	de:l(a)
EgCai	da	di	do:l	
IqAq	hada:	hada:~ha(:)dya:	hawda:~hawde:	
IqBgG	ha:ða	ha:ði ha:y	haðo:la	(ha)ðanni
IqBgJ	ha:ða	ha:yi	haðo:li	
IqKhz	ha:ð(a)	ha:ði~ha:y	(ha)ðo:l(a)	(ha)ðanni
IqMosJ	ha:ða	ha:yi	haðo:li	
LvBsh	ha:da~ha:d~ha:	haydi~hayd~hay	hawdi~hawd~haw	
LvDam	ha:da	ha:di hayy	hado:l	

LvHeb	ha:da	ha:di	hado:l(a)	
LvS101	ha:da	hayye	haw(w)	
LvS239	ha:da	ha:di	hawdi	
LvS281	ha:ða	ha:ði	haðo:l	haðanna
LvS305	ha:da	he:di	haðo:l	
LvS330	ha:ða	ha:ði	haðo:l	
LvS340	he:ða	ha:y	haðo:la	haðanna
LvS348	he:ð	ha:y	hawð(i)	
LvS417	ha:ða	hay	haðo:l	
LvS433	ha:ð ^ʰ a	ha:y	haðo:l(a)	
LvSPal	he:ðya~he:ðe	he:ðei	haðo:l	
LvSSou	ha:ði	ha:ði	(ha)ðol	
Malt	da ~dan(a)	di ~din(a)	daw ~dawn(a)	
NaBen	'haðɑ ('hað ^ʰ ɑ)	haði	ha'ð ^ʰ o:l	ha'ðe:n
NaChr	ha:ða	ha:ði	ha:ðu	
NaDj	ha:da	ha:di	ha:du	
NaFez	ha:da	ha:di	ha:du	
NaHas	(ha:)ðæ	(ha:)ði	(ha:)ðu	
NaMz	ha:da	ha:di	ha:du	
NaTl	ha:da	ha:di	ha:du	
NaTrJ	ada	adi	ad ^ʰ un(i)	
NaTrM	ha:da(:ya)~a:hw ^ʰ á	ha:di(:ya)~a:hyá	ha:du(:ma(:ya))~a:h(u)m ^ʰ á	
NaTun	haða (aða) haðaya	haði (aði) haðiya	haðuma~haðum~haðula	
Nig	'da	'di	'do:l(a)	'de:l(a)
Shuk	da	di	do:l	de:l
SinR	ha:ða	ha:ðiy~he:ðiy	ha:ðo:l(ah)~ho:ð ^ʰ al ^ʰ (l ^ʰ ah)	
SinS	ha:ða	ha:ðiy	ha:ðo:l(ah)~ho:ð ^ʰ al ^ʰ (l ^ʰ ah)	

UzbA	ha:d~hat~ha:z	ha:di:	hadla:u	hadla:n
UzbDj	ha:d~hat~ha:z	hai	halo:~hala:u	hala:n
Y104	ha:ða	ha:ði	ðawla~ðo:la	
Y121	ha:ða	ha:ði	hawla~ha:ðawla~ha:ðo:la	
Y145	ha:ða	ha:ði	ha:ðum	ha:ðe:n
Y156	ðe:	ði:	ðinne:n	
Y24	ha:ða	ha:ði	hawla	
Y6	ðiyya	tiyya	?o:li:	?e:li:
Y96	ða	ta	ðayl~ðe:la	
Y99	ha:ða	ha:ði	?	
YHadr	ða	ði	ðe:l(a) ðo:l(a)	
YTalh	ða:	ðih	?awla:	
YWas	ha:ða	ha:ði	haðowla	
YXash	ða:	ta:	?o:ða:	

Table A.2: Demonstrative data from sample: Distal forms

Dialect name	M.SG	F.SG	M./C. PL	F. PL
AfgA				
And	'haðak~'ðik		hawlak	
AnDar	a:k	ayk(i)	o:k	
ArAnz	(ha)ða:k	(ha)ði:tʃ	(ha)ðela:k	
ArBah	(ha(:))da:k~ha:k(a)	(ha(:))di:k~he:k(a~i)	(ha(:))dela:k~hado:la:k	
ArEm	(ha)ða:k	(ha)ði:tʃ	(ha)ðo:la:k	(ha)ðe:la:k~(ha)ðila:k
ArNjd	(ha)ða:k	(ha:)di:tʃ	(ha)ðo:la:k	(ha:)ðo:litʃ
ArOm	(ha:)ða:k	(ha:)dik	(ha:)ðe:la:k	
ArQah	ða:k	ta:k	ʔawða:k	
ArQz	ða:kak	ðikah	ðalla:kah	
Chd	dak	dik	dolak	
ChdSh	da:k	di:k	do:la(:)k	de:la(:)k
Cyp	aðak~ak	aðik~ayk	allik	
EgBr	dukka:ti	dikki:ti	dukkum ^ʔ m ^ʔ a	dikkinna
EgCai	dukha	dikha	dukham~dukhum~dukhumma	
IqAq	ha:k~hada:k	ha:k~hada:k	hawk~hawda:k	
IqBgG	(ha)ða:k(a)	(ha)ði:tʃ	(ha)ðo:la:k	
IqBgJ	haða:k	haði:k	haðo:lak	
IqKhz	ða:k(a)	(ha:)ði:tʃ	ðo:la:(a)	(ha:)ðannitʃ
IqMosJ	haða:k(a)	haði:k(a)	haðo:k(a)	
LvBsh	hida:k(i)	hidi:k(i)~haydi:k(i)	hudi:k(i)~hawdi:k(i)	
LvDam	hada:k	hadi:k	hado:li:k hadənk	
LvHeb	hada:k	hadi:k(e)	hado:la:k(e)	
LvS101	hada:k	hadi:k	hawki	

LvS239	ha:ka	ha:ki	hawke:n	
LvS281	haða:k	haði:tʃ	haðo:la:k	haðannitʃ
LvS305	hada:k	hadi:k	hawwi:k	
LvS330	haðe:k	haði:ke	hado:k	
LvS340	haðe:k	haði:ki	hadawle:k	
LvS348	hede:k	heði:ki	hawði:ki	
LvS417	haða:k	haði:k	?	
LvS433	haða:k	haði:tʃ	haðo:la:k	
LvSPal	(ha)ðe:k	(ha)ði:k	(ha)ðo:le:k~(ha)ðu:le:k	
LvSSou	haða:tʃ	haði:tʃ	(ha)ðo:latʃ	
Malt	dak(a)	dik(a)	dawk(a)	
NaBen	ha'ðʰak(ki)	ha'ðik(ki)		
NaChr	ha:ða:k	ha:ði:k	ha:ðu:k	
NaDj	ha:da:k	ha:di:k	ha:du:k	
NaFez	ha:da:k	ha:di:k	ha:du:k	
NaHas	ða:k	ði:k	ðu:k	
NaMz	(ha:)ða:ʰk	(ha:)di:k	(ha:)ðu:k~(ha:)du:m	
NaTl	ha:dak	ha:dik	ha:duk	
NaTrJ	(a)dak	(a)dik	(a)dʰuk	
NaTrM	ha:da:ka(:y(a))	ha:di:ka(:y(a))	ha:du:k-a~a:y~a:ya~umma	
NaTun	(h)aðaka~haka	(h)adika	hadukum~hadumka	
Nig	'dʰa:ka	'dʰi:ka	'dʰo:lak(a)	'dʰo:lak(a)
Shuk	da:k	di:k	do:lak	de:lak
SinR	ha:ða:k	he:ði:k	haðʰo:lʰa:k	
SinS	ha:ðʰa:k	ha:ði:kih	ha:ðʰalʰa:k~ho:ðʰalʰa:k	
UzbA	du:k~duk	duki:	dukala:u	dukala:n
UzbDj	du:k~duk	diki:	dukala:	dikala:n~dikila:n

Y104	ha:ða:k	ha:ta:k	ðawla:k~ðo:la:k	
Y121	ha:ða:k	ha:ði:k	ðawla:k~ðo:la:k	
Y145	ha:ða:k	ha:ði:k	ha:ðunk	ha:ðink
Y156	ð ^ʰ a:k	ð ^ʰ e:k	ðinne:k	
Y24	ha:ða:k	ha:ði:k	hawla:k~ho:la:k	
Y6	ða:k	ta:k	ʔo:lak	ʔe:lak
Y96	ða:k	ta:k	ðe:lak	
Y99	ha:ða:k	ha:ði:k	haw ^ʰ a:k	
YHadr	ða:k	ði:k	ðe:la:k	
YTalh	ða:k	ta:k		
YWas	ha:ða:k	ha:ði:k		
YXash	ða:k	ta:k		

Bibliography

- Abboud-Haggar, Soha (2006). "Dialects: Genesis". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Vol. 1. Leiden: Brill, pp. 613–622.
- Abu-Haidar, Jareer (1988). "The language and function of the Hispano-Arabic Zǧal". In: *A Miscellany of Middle Eastern Articles: In Memoriam Thomas Muir Johnstone*. Essex: Longman.
- Abu-Mansour, Mahasen Hasan (2006). "Meccan Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 181–185.
- Abun-Nasr, Jamil M. (1987). *A history of the Maghrib in the Islamic Period*. Cambridge: Cambridge University Press.
- Adamou, Evangelia (2012). "Verb morphologies in contact: evidence from the Balkan area". In: *Morphologies in contact*. Ed. by M. Vanhove et al. Berlin: Akademie Verlag.
- (forthcoming). "Change and Variation in a Trilingual Setting: Evidentiality in Pomak". In: *The interplay of variation and change in contact settings – Morphosyntactic studies*. Ed. by Léglise I. and C. Chamoreau. John Benjamins Publishing Company.
- Adams, R.M.C. (1965). *Land behind Baghdad: a history of settlement on the Diyala Plains*. University of Chicago Press.
- Agency, United States Central Intelligence (1973). *Atlas [of] issues in the Middle East*. Washington, D.C.
- Agius, Dionisius A. (1996). *Siculo Arabic*. New York: Kegan Paul International.
- Aikhenvald, Alexandra Y. (2003). *Language Contact in Amazonia*. Oxford: Oxford University Press.
- (2004). *Evidentiality*. Oxford: Oxford University Press.

- Ajaya, J.F. Ade and Michael Crowder, eds. (1985). *Historical Atlas Of Africa*. New York: University of Cambridge Press.
- Andersen, Henning (1988). "Centre and periphery: adoption, diffusion and spread". In: *Historical Dialectology*. Ed. by Jacek Fisiak. Berlin: Mouton de Gruyter.
- (1996). *Reconstructing Prehistorical Dialects: Initial Vowels in Slavic and Baltic*. Mouton de Gruyter.
- Anderson, S.R. and E.L. Keenan (1985). "Deixis". In: *Language Typology and Syntactic Distinction*. Ed. by T. Shopen. Cambridge: Cambridge University Press.
- al-Ansary, A.R. (1982). *Qaryat al-Fau, a portrait of pre-Islamic civilization in Saudi Arabia*. Riyadh: University of Riyadh.
- Ashtor, Eliyahu et al. (2007). "Spain". In: *Encyclopaedia Judaica*. Ed. by Michael Berenbaum and Fred Skolnik. 2nd. Vol. 19. Detroit: Macmillan.
- Auer, Peter (2002). "The Construction of Linguistic Borders and the Linguistic Construction of Borders". In: *Dialects Across Borders: Selected papers from the 11th international conference on methods in dialectology (Methods XI) Joensuu, August 2002*. Ed. by Markku Filppula et al. Amsterdam: John Benjamins Publishing Company.
- Al-Azraqi, Munira (2010). "The Ancient Dad in Southwest Saudi Arabia". In: *Arabica* 57, pp. 57–67.
- Bailey, G., T. Wikle, and L. Sand (1993). "Some patterns of linguistic diffusion". In: *Language Variation and Change* 5, pp. 359–90.
- Barfield, Thomas J. (1993). *The nomadic alternative*. Englewood Cliffs, N.J.: Prentice Hall.
- Barth, Frederik (1961). *Nomads of South Persia*. Boston: Little, Brown.
- Barth, Jakob (1913). *Die Pronaminalbildung in den semitischen Sprachen*. Leipzig: J.C. Hinrichs'sche Buchhandlung.
- Bashear, Suliman (1989). "Yemen in Early Islam: An Examination of Non-Tribal Traditions". In: *Arabica* 36.
- (1997). *Arabs and Others in early Islam*. Princeton, NJ: Darwin Press.
- Batibo, Herman (1992). "The Fate of Ethnic Languages in Tanzania". In: *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*. Ed. by Matthias Brenzinger. Berlin: Mouton de Gruyter.
- Behnstedt, Peter (1985). *Die nordjemenitischen Dialekte*. Wiesbaden: L. Reichert.
- (1987). *Die Dialekte der Gegend von Sa'dah (Nord-Jemen)*. Wiesbaden: Otto Harrassowitz.
- (1994). *Der arabische Dialekt von Soukhne (Syrien)*. *Semitica viva*. Harassowitz.

- Behnstedt, Peter (1997). *Sprachatlas von Syrien*. Weisbaden: Harrassowitz.
- Behnstedt, Peter and Manfred Woidich (1985). *Die ägyptisch-arabischen Dialekte*. In 5 volumes. Weisbaden: L. Reichert.
- (2005). *Arabische Dialektgeographie: eine Einführung*. Boston: Brill.
- Beier, Christine, Lev Michael, and Joel Sherzer (2002). “Discourse forms and processes in indigenous lowland South America: an areal-typological perspective”. In: *Annual Review of Anthropology* 31, pp. 121–45.
- Beyer, Klaus (1986). *The Aramaic Language: Its Distribution and Subdivisions*. Translated from the German by John P. Healey. Göttingen, Germany: Vandenhoeck and Ruprecht.
- Blanc, Haim (1953). *Studies in North Palestinian Arabic; linguistic inquiries among the Druzes of western Galilee and Mt. Carmel*. Jerusalem: Israel Oriental Society.
- (1964). *Communal dialects in Baghdad*. Cambridge, MA: Harvard University Press.
- Bloch, Ariel A (1986). “Presentative Structures and their Syntactic and Semantic Development”. In: *Studies in Arabic Syntax and Semantics*. Wiesbaden: Otto Harrassowitz.
- Bloomfield, Leonard (1933). *Language*. New York: Holt.
- Borg, Alexander (1985). *Cypriot Arabic*. Stuttgart: Kommissionsverlag Franz Steiner Wiesbaden GMBH.
- (2004). *A Comparative Glossary of Cypriot Maronite Arabic: With an Introductory Essay (Handbook of Oriental Studies)*. Brill Academic Pub.
- (2006). “Cypriot Maronite Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 536–543.
- Bosch, E. et al. (2000). “Genetic structure of north-west Africa revealed by STR analysis”. In: *European Journal of Human Genetics* 8.5, pp. 360–366.
- Bosworth, C. E. (1983). “Iran and the Arabs before Islam”. In: *The Cambridge History of Iran*. Ed. by Alan Bowman, Averil Cameron, and Peter Garnsey. Cambridge University Press.
- Bowern, Claire (2006). “Another look at Australia as a linguistic area”. In: *Linguistic Areas: convergence in historical and typological perspectives*. New York: Palgrave Macmillan.
- Bowern, Claire et al. (2011). “Does lateral transmission obscure inheritance in hunter-gatherer languages?” In: *PloS one* 6.9, e25195.

- Brincat, Joseph M. (2011). *Maltese and Other Languages: A Linguistic History of Malta (Maltese Social Studies)*. Midsea Books.
- Britain, David (2008). "Space and Spatial Diffusion". In: *The Handbook of Language Variation and Change*. Ed. by J.K. Chambers, Peter Trudgill, and Natalie Schilling-Estes. Blackwell Publishing Ltd, pp. 603–637.
- Britain, David and Peter Trudgill (2005). "New Dialect Formation and Contact-induced Reallocation: Three Case Studies from the English Fens". In: *International Journal of English Studies* 5.1, pp. 183–209.
- Brockelmann, Carl (1908). *Grundriss der vergleichenden Grammatik der semitischen Sprachen*. Berlin: Verlag Von Reuther and Reichard.
- (1928). "Deminutiv und Augmentativ im Semitischen". In: *Zeitschrift für Semitistik und verwandte Gebiete* 6, pp. 109–133.
- Brown, Jonathan A.C. (2007). "New Data On The Delateralization Of Dad And Its Merger With Za' In Classical Arabic: Contributions From Old South Arabic And The Earliest Islamic Texts On D / Z Minimal Pairs". In: *Journal of Semitic Studies* 52.2, pp. 335–368.
- Brown, Penelope and Stephen C. Levinson (1979). "Social Structure, groups and interaction". In: *Social markers in speech*. Ed. by Klaus R. Scherer and Howard Giles. Cambridge: Cambridge University Press.
- Brustad, Kristen (2000). *The Syntax of Spoken Arabic: a comparative study of Moroccan, Egyptian, Syrian and Kuwaiti Dialects*. Washington, D.C.: Georgetown University Press.
- (2008). "Drink Your Milks! āt as Individuation Marker in Levantine Arabic". In: *Classical Arabic Humanities in Their Own Terms: Festschrift for Wolfhart Heinrichs on his 65th Birthday Presented by his Students and Colleagues*. Ed. by B. Gruendler and M. Cooperson. Leiden: Brill, pp. 1–19.
- Buchheit, Robert H. (1988). "Language Shift in the concentrated Mennonite district of Kansas". In: *International Journal of the Sociology of Language* 69.1, pp. 5–18.
- Bühler, Karl (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer: Fischer.
- Bulliet, Richard W. (1979). *Conversion to Islam in the medieval period : an essay in quantitative history*. Cambridge, MA: Harvard University Press.
- Bybee, Joan (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bynon, Theodora (1977). *Historical Linguistics*. Cambridge: Cambridge University Press.

- Campbell, Lyle (2004). *Historical Linguistics: An Introduction*. Second edition. The MIT Press.
- (2006). “Area Linguistics: A Closer Scrutiny”. In: *Linguistic Areas: Convergence in Historical and Typological Perspective*. Ed. by Yaron Matras, April McMahon, and Nigel Vincent. Palgrave Macmillan.
- Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark (1986). “Meso-America as a Linguistic Area”. English. In: *Language* 62.3, pp. 530–570.
- Cantineau, Jean (1934). *Le Dialecte Arabe De Palmyre*. Vol. 1. Beirut: Mémoires de L’Institut Français de Damas.
- Caskel, W. (2012a). “al-‘Arab (ii): The expansion of the Arabs: General, and the “Fertile Crescent””. In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- (2012b). “Bakr b. Wā’il”. In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- (2012c). “‘Akk”. In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Caubet, Dominique (1993). *L’Arabe Marocain: Phonologie et Morphosyntaxe*. Paris-Louvain: Éditions Peeters.
- Cherni, L. et al. (2005). “Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa”. In: *Human biology* 77.1, pp. 61–70.
- Clements, J. Clancy (2009). *The Linguistic Legacy of Spanish and Portuguese: Colonial Expansion and Language Change*. Cambridge: Cambridge University Press.
- Cohen, Anthony Paul (1985). *The Symbolic Construction of Community*. Chichester: E. Horwood.
- Cohen, David (1970). “Les deux parlers arabes de Tunis”. In: *Études de linguistique sémitique et arabe*. The Hague and Paris: Mouton, 150–171.
- Cohen, David and Mohammed el Chennafi (1963). *Le dialecte arabe ḥassānīya de Mauritanie, parler de la Gabla*. Paris: C. Klincksieck.
- Cohen, Mark R. (1999). “What was the Pact of ‘Umar? A literary-historical study”. In: *Jerusalem Studies in Arabic and Islam* 23, pp. 100–151.
- Conde-Silvestre, Juan Camilo (2012). “The Role of Social Networks and Mobility in Diachronic Sociolinguistics”. In: *The Handbook of Historical Sociolinguistics*. Ed. by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre. Malden, MA: Wiley-Blackwell.
- Conrad, Lawrence I. (1981). “The Plague in the Early Medieval Near East”. PhD thesis. Princeton University.

- Corcos, David, Hayyim J. Cohen, and Michael M. Laskier (2007). "Morocco". In: *Encyclopaedia Judaica*. Ed. by Michael Berenbaum and Fred Skolnik. 2nd. Vol. 14. Detroit: Macmillan Reference USA, pp. 493–507.
- Corcos, David, Haim Saadoun, and Robert Attal (2007). "Algeria". In: *Encyclopaedia Judaica*. Ed. by Michael Berenbaum and Fred Skolnik. 2nd. Vol. 1. Detroit: Macmillan Reference USA, pp. 645–652.
- Corriente, Federico C. (1977). *A grammatical sketch of the Spanish Arabic dialect bundle*. Madrid: Instituto Hispano-Arabe de Cultura.
- Cowan, William George (1960). "A Reconstruction of Proto-Colloquial Arabic". PhD thesis. Cornell University.
- Cowell, Mark W. (1964). *A Reference Grammar of Syrian Arabic, based on the dialect of Damascus*. Washington, DC: Georgetown University Press.
- Cremona, Joseph (1994). "The survival of Arabic in Malta: the Sicilian centuries". In: *The changing voices of Europe : social and political changes and their linguistic repercussions, past, present and future*. Cardiff: University of Wales Press.
- Croft, William (2000). *Explaining Language Change*. Pearson Education ESL.
- (2003). "Social evolution and language change". Unpublished manuscript.
- Crone, Patricia (2012). "Mawlā". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Crystal, David (1997). *The Cambridge Encyclopedia of the English Language*. Oxford: Oxford University Press.
- Dahl, Östen (2001). "Principles of Area Typology". In: *Language Typology and Language Universals*. Ed. by Martin Haspelmath et al. Berlin: Mouton de Gruyter, pp. 1456–1470.
- Davies, Humphrey Taman (1981). "17th-century Egyptian Arabic: A Profile Of The Colloquial Material In Yūsuf al-Širbīnī's Hazz al-Quhūf Fī Šarḥ Qaṣīd Abī Šādūf". Ph.D. University of California, Berkeley.
- Diem, Werner (1973). *Skizzen jemenitischer Dialekte*. Wiesbaden: Verlag.
- Diessel, Holger (1997). "Predicative Demonstratives". In: *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, pp. 72–82.
- (1999). *Demonstratives : form, function, and grammaticalization*. Amsterdam: John Benjamins Publishing Company.
- (2012). "Bühler's two-field theory of pointing and naming and the deictic origins of grammatical morphemes". In: *Grammaticalization and Language*

- Change: New reflections*. Ed. by Kristin Davidse et al. John Benjamins Publishing Company.
- Dillard, J. L. (1972). *Black English: Its history and usage in the United States*. New York: Random House.
- Dixon, R. M. W. (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- (2003). “Demonstratives: A cross-linguistic typology”. In: *Studies in Language* 27.1, pp. 61–112.
- Djaït, Hichem (2007). “Kufa (al-Kūfa)”. In: *Historic cities of the Islamic world*. Ed. by C. Edmund Bosworth. Leiden: Brill.
- Donner, Fred McGraw (1981). *The Early Islamic Conquests*. Princeton, New Jersey: Princeton University Press.
- (1989). “The Role of Nomads in the Near East in Late Antiquity (400–800 C.E.)”. In: *Tradition and Innovation in Late Antiquity*. Ed. by F.M. Clover and R.S. Humphreys. Madison: University of Wisconsin Press.
- Donohue, Mark, Søren Wichmann, and Mihai Albu (2008). “Typology, Areality, and Diffusion”. In: *Oceanic Linguistics* 47, pp. 223–232.
- Donzel, E. van (2012). “Sirhān”. In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Doss, Madiha (1979). “The position of the demonstrative da, di in Egyptian Arabic: A diachronic inquiry.” In: *Annales Islamologiques* 15, pp. 349–357.
- Dr. Ludwig Reichert Verlag (1992). “Vorderer Orient: Nomadismus und anderer Formen der Wanderviehwirtschaft [Middle East: Nomadism and other Forms of Pastoral Migration]”. In: *Tübinger Atlas des Vorderen Orients (TAVO)*. Map. Wiesbaden: Verlag.
- Dresch, P. (1989). *Tribes, Government, and History in Yemen*. Clarendon Paperbacks Series. Clarendon Press.
- Dryer, Matthew S. (1989). “Large Linguistic Areas and Language Sampling”. In: *Studies in Language* 13.2, pp. 257–292.
- Dumper, Michael R. T. and Bruce E Stanley (2006). *Cities of the Middle East and North Africa : A Historical Encyclopedia*. Santa Barbara: ABC-CLIO.
- Duri, A. A. (2012). “Bagḥdād”. In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Dyen, Isidore (1969). “Reconstruction, the Comparative Method, and the Proto-Language Uniformity Assumption”. English. In: *Language* 45.3, pp. 499–518.

- Edwards, John (2009). *Language and Identity: an introduction*. Cambridge: Cambridge University Press.
- Edzard, Lutz (2006). "Omani Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 478–491.
- E.Kogan, Leonid and Andrey V.Korotayev (1997). "Sayhadic (Epigraphic South Arabian)". In: *The Semitic Languages*. Ed. by Robert. Routledge.
- El'Ad, Amikam (1982). "The Coastal Cities of Palestine during the Early Middle Ages". In: *The Jerusalem Cathedra: Studies in the History, Archaeology, Geography and Ethnography of the Land of Israel*. Vol. Volume II. Detroit: Wayne State University Press, pp. 146–167.
- Elisséeff, N. (2012). "Mardj Rāhiṭ". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Enfield, N. J. (2003). *Linguistic Epidemiology*. Routledge Curzon.
- Epps, Patience (2005). "Areal diffusion and the development of evidentiality: Evidence from Hup." In: *Studies in Language* 29.3, pp. 617–650.
- Escure, Genevieve (1997). *Creole and Dialect Continua: Standard Acquisition Processes in Belize and China*. Amsterdam: Benjamins.
- Faber, Alice (1997). "Genetic Subgrouping of the Semitic Languages". In: *The Semitic Languages*. Ed. by Robert Hetzron. New York: Routledge.
- Fagyal, Zsuzsanna et al. (2010). "Centers and peripheries: Network roles in language change". In: *Lingua* 120.8. Asymmetries in Language Acquisition, pp. 2061–2079.
- Fares, Saba (2003). "La chronologie des inscriptions dédanites et lihyânites d'al-'Ulâ : état de la question". In: *Topoi Orient et Occident* suppl. 4, pp. 379–405.
- al-Farrā', Abi Zakariyyā Yaḥyā bin Ziyād ([d. 822] 1983). *Ma'āniy al-Qur'ān [Meanings of the Quran]*. Ed. by Ahmad Yūsif Nağātī and Muḥammad 'Alī al-Nağğār. 3rd edition. Beirut: 'Ālam al-Kutub.
- Ferguson, Charles A. (1959). "The Arabic Koine". In: *Language* 35, pp. 616–630.
- Finkelstein, Israel and Avi Perevolotsky (1990). "Processes of Sedentarization and Nomadization in the History of Sinai and the Negev". In: *Bulletin of the American Schools of Oriental Research* 279, pp. 67–88.
- Fischer, Greg (2004). "A New Perspective on Rome's Desert Frontier". In: *Bulletin of the American Schools of Oriental Research* 336, pp. 49–60.
- (2011). *Between Empires: Arabs, Romans, and Sasanians in Late Antiquity*. Oxford: Oxford University Press.

- Fischer, Wolfdietrich (1959). *Die demonstrativen Bildungen der neuarabischen Dialekte: Ein Beitrag zur historischen Grammatik des Arabischen*. The Hague: Mouton & Co.
- (2002). *A Grammar of Classical Arabic: Third Revised Edition*. Translated from the German by Jonathan Rogers. Yale University Press.
- Fischer, Wolfdietrich and Otto Jastrow (1980). *Handbuch der arabischen Dialekte*. Weisbaden: Otto Harrasowitz.
- Foley, John Miles (1996). "Guslar and Aoidos: Traditional Register in South Slavic and Homeric Epic". In: *Transactions of the American Philological Association* (1974-) 126, pp. 11–41.
- Fought, Carmen (2012). "Ethnic Identity and Linguistic Contact". In: *The Handbook of Language Contact*. Ed. by Raymond Hickey. Malden, M.A.: Wiley-Blackwell.
- Fück, Johann (1950). *Arabiya, Untersuchungen zur arabischen Sprach- und Stilgeschichte*. Berlin: Akademie-Verlag.
- Gaibar, Maria et al. (2011). "Genetic differences among North African Berber and Arab-speaking populations revealed by Y-STR diversity." In: *Annals of Human Biology* 38.2, pp. 228–236.
- Garbell, Irene (1958). "Remarks on the Historical Phonology of An Eastern Mediterranean Arabic Dialect". In: *Word* 14, pp. 303–337.
- Gérard, N. et al. (2006). "North African Berber and Arab influences in the western Mediterranean revealed by Y-chromosome DNA haplotypes". In: *Human biology* 78.3, pp. 307–316.
- George, A. R. (1973). "Processes Of Sedentarization Of Nomads In Egypt, Israel And Syria: A Comparison". In: *Geography* 58.2, pp. 167–169.
- Gibson, Maik (2006). "Tunis Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 563–571.
- Glick, Thomas F. (1979). *Islamic and Christian Spain in the Early Middle Ages*. Princeton University Press.
- Goossens, Jan (1977). *Deutsche Dialektologie*. Sammlung Götschen. Berlin: Walter de Gruyter.
- Grace, George W. (1996). "Regularity of Change in What?" In: *The Comparative Method Reviewed : Regularity and Irregularity in Language Change*. Ed. by Malcolm Ross and Mark Durie. Oxford University Press.
- Grand'Henry, Jacques. (1972). *Le parler arabe de Cherchell, Algérie*. Louvain-la-Neuve: Université catholique de Louvain, Institut orientaliste.

- Grand'Henry, Jacques (1976). *Les Parlers Arabes de la Région du Mzāb (Sahara Algérien)*. Leiden: E.J. Brill.
- Grassi, Vincenza (2012). "Şikilliya". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Greenberg, Joseph H. (1960). "An Afro-Asiatic Pattern of Gender and Number Agreement". English. In: *Journal of the American Oriental Society* 80.4, pp. 317–321.
- Griffith, Sidney H. (2008). *The church in the shadow of the mosque : Christians and Muslims in the world of Islam*. Princeton, NJ: Princeton University Press.
- Gumperz, John J. and Robert Wilson (1971). "Convergence and Creolization : A Case From the Indo-Aryan/Dravidian Border in India". In: *Pidginization and Creolization of Languages*. Ed. by Dell Hymes. Cambridge: Cambridge University Press.
- Gundel, Jeanette K., Mamadou Bassene, et al. (2010). "Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation". In: *Journal of Pragmatics* 42.7, pp. 1770–1785.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski (1993). "Cognitive Status and the Form of Referring Expressions in Discourse". English. In: *Language* 69.2, pp. 274–307.
- Hagège, Claude (1973). *Profil d'un parler arabe du Tchad*. Paris: P. Geuthner.
- al-Ḥamrūnī, Aḥmad (2009). *al-Hijrāt al-Andalusīyah ila al-bilād al-Tūnisīyah [Andalusian Migrations to Greater Tunisia]*. Tunis: Mīdyākūm.
- Harris, William (2012). *Lebanon: A History, 600-2011 (Studies in Middle Eastern History)*. Oxford University Press.
- Harvey, L. P. (1990). *Islamic Spain, 1250 to 1500*. University Of Chicago Press.
- (2005). *Muslims in Spain, 1500 to 1614*. University Of Chicago Press.
- Haspelmath, Martin (2006). "Against markedness (and what to replace it with)". In: *Journal of Linguistics* 42.1, pp. 25–70.
- Hasselbach, Rebecca (2007a). "Demonstratives in Semitic". English. In: *Journal of the American Oriental Society* 127.1, pp. 1–27.
- (2007b). "External Plural Markers in Semitic: A New Assessment". In: *Studies in Semitic and Afroasiatic Linguistics Presented to Gene B. Gragg*. Ed. by Cynthia L. Miller. Chicago: The Oriental Institute of the University of Chicago.
- Hazen, Kirk (2000). *Identity and ethnicity in the rural South: A sociolinguistic view through past and present*. Publications of the American Dialect Society 83. Durham, NC: Duke University Press.

- Heath, Jeffrey (2002). *Jewish and Muslim Dialects of Moroccan Arabic*. Routledge-Curzon Arabic Linguistics Series. RoutledgeCurzon.
- Henderson, Anita (1996). "The short a pattern of Philadelphia among African American speakers". In: *University of Pennsylvania Working Papers in Linguistics* 3, pp. 127–40.
- Hendery, Rachel (2012). *Relative Clauses in Time and Space : A case study in the methods of diachronic typology*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Henkin, Roni (2010). *Negev Arabic: Dialectal, Sociolinguistic and Stylistic Variation*. Wiesbaden: Harrassowitz Verlag.
- Hernández, Miguel Cruz (1998). "The settlement and organisation of the Syrian Junds in al-Andalus". In: *The Formation of al-Andalus, Part 1: History and Society*. Ed. by Manuela Marín. Ashgate Variorum.
- Hernández-Campoy, Juan Manuel and Juan Camilo Conde-Silvestre, eds. (2012). *The Handbook of Historical Sociolinguistics*. Wiley-Blackwell.
- Hetzron, Robert (1976). "Two Principles of Genetic Reconstruction". In: *Lingua* 38, pp. 89–108.
- Himmelman, Nikolaus P. (1996). "Demonstratives in Narrative Discourse". In: *Studies in Anaphora*. Ed. by Barbara Fox. Amsterdam: John Benjamins Publishing Company.
- Hinds, Martin and El-Said Badawi (1986). *A Dictionary of Egyptian Arabic: Arabic-English*. Beirut, Lebanon: Librarie Du Liban.
- Hock, Hans Henrich (1991). *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Hockett, Charles F. (1958). *A course in modern linguistics*. New York: Macmillan.
- Holes (1987). *Language Variation And Change In A Modernising Arab State: The Case Of Bahrain*. Library of Arabic Linguistics, Monograph 7. London: Routledge.
- Holes, Clive (2006). "The Arabic Dialects of Arabia". In: *Proceedings of the Seminar for Arabian Studies, Vol. 36, Papers from the thirty-ninth meeting of the Seminar for Arabian Studies held in London, 21-23 July 2005*. Archaeopress.
- Hopkins, Simon (1984). *Studies in the Grammar of Early Arabic. Based upon papyri datable to before 300 A.H./912 A.D.* Oxford University Press.
- Horrocks, Geoffrey (1997). "Homer's Dialect". In: *A New Companion to Homer*. Ed. by Brian Powell and Ian Morris. Leiden: Brill. Chap. 2.

- Hourani, Guita G. (1998). "A Reading in the History of the Maronites of Cyprus From the Eighth Century to the Beginning of British Rule". In: *Journal of Maronite Studies*.
- Hoyland, Robert G. (2001). *Arabia and the Arabs: From the Bronze Age to the Coming of Islam*. New York: Routledge.
- (2004). "Language and Identity: The Twin Histories of Arabic and Aramaic (and: Why did Aramaic Succeed where Greek Failed?)" In: *Scripta Classica Israelica* 23, pp. 183–199.
- (2009). "Arab kings, Arab tribes and the beginnings of Arab historical memory in late Roman epigraphy". In: *From Hellenism to Islam: Cultural and Linguistic Change in the Roman Near East*. Ed. by Hannah M. Cotton et al. Cambridge: Cambridge University Press.
- Huehnergard, John (1983). "Asseverative *la and Hypothetical *lu/law in Semitic". In: *Journal of the American Oriental Society* 103.3, pp. 569–593.
- (2005). "Features of Central Semitic". In: *Biblical and Oriental Essays in Memory of William L. Moran*. Ed. by Agustinus Gianto. Rome: Pontificio Istituto Biblico.
- (2006). "On the Etymology of the Hebrew Relative se". In: *iblical Hebrew in Its Northwest Semitic Setting. Typological and Historical Perspectives*. Ed. by S.E. Fassberg and A. Hurvitz. Winona Lake: Eisenbrauns, pp. 103–125.
- (2011). "Phyla and Waves: Models of Classification of the Semitic Languages". In: *The Semitic Languages: An International Handbook*. Ed. by Stefan Weninger et al. De Gruyter Mouton.
- (2012). "Arabic as a Semitic Language". Unpublished paper presented as a keynote speech at the Arabic Linguistics Symposium, April 9, 2010 at the University of Texas at Austin.
- Huehnergard, John and Na'ama Pat-El (2007). "Some Aspects of Cleft in Semitic Languages". In: *Studies in Semitic and General Linguistics in Honor of Gideon Goldenberg*. Ed. by Tali Bar and Eran Cohen. Münster: Ugarit-Verlag.
- (2012). "Third-person possessive suffixes as definite articles in Semitic". In: *Journal of Historical Linguistics* 2:1 2.1, pp. 25–51.
- Ibn Khaldūn (1377 CE / 2006). *al-Muqaddama*. Ed. by Darwish al-Juweidī. al-Maktaba al-Aṣriyya.
- Imai, Shingo (2003). "Spatial deixis". Unpublished Ph.D. thesis.
- Ingham, B. (1982). "Notes on the dialect of the Dhafir of north-eastern Arabia". In: *BSOAS*, 45, ii, 249–297.

- Ingham, Bruce (1971). "Some Characteristics of Meccan Speech". English. In: *Bulletin of the School of Oriental and African Studies, University of London* 34.2, pp. 273–297.
- (1973). "Urban and Rural Arabic in Khūzistān". English. In: *Bulletin of the School of Oriental and African Studies, University of London* 36.3, pp. 533–553.
- (1986). *Bedouin Of Northern Arabia: Traditions Of The Āl-Dafīr*. London: KPI.
- (2006a). "Afghanistan Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 28–35.
- (2006b). "Khuzestan Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 2. Leiden: Brill, pp. 571–578.
- (2006c). "Najdi Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 326–334.
- Ionin, Tania (2006). "This is definitely specific: Specificity and Definiteness in Article Systems". In: *Natural Language Semantics* 14, pp. 175–234.
- al-Iṣfahānī, Abū l-Faraj (1969-1982). *Kitāb al-aghānī*. Ed. by Ibrāhīm al-Abyārī. Cairo: Dār al-Shaʿb.
- Islamic Studies of the University of Zaragoza, Institute of, ed. (2012). *A Descriptive and Comparative Grammar of Andalusī Arabic*. Handbook of Oriental Studies. Brill Academic Publishers.
- Al-Jallad, Ahmad (2009). "The Polygenesis of the Neo-Arabic Dialects". In: *Journal of Semitic Studies* 54, pp. 515–536.
- (2012). "Ancient Levantine Arabic: A Reconstruction Based on the Earliest Sources and the Modern Dialects". PhD thesis. Harvard University.
- Jarbou, Samir Omar (2010). "Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic". In: *Journal of Pragmatics* 42.11, pp. 3078–3097.
- (2012). "Medial Deictic Demonstratives In Arabic: Fact Or Fallacy". In: *Pragmatics* 22.1, pp. 103–118.
- Jastrow, Otto (1973). *Daragözü: eine arabische Mundart der Kozluk-Sason-Gruppe (Südostanatolien)*. *Grammatik und Texte*. Nürnberg: Verlag Hans Carl.
- (1978). *Die Mesopotamisch-Arabischen Qeltu-dialekte. Band I: Phonology und Morphologie*. Wiesbaden: Kommissionsverlag Franz Steiner GMBH.
- (1979). "Zur Arabischen Mundart Von Mossul". In: *Zeitschrift für Arabische Linguistik* 2, pp. 36–75.

- Jastrow, Otto (1990). *Der arabische Dialekt der Juden von Agra und Arbil*. Weisbaden: O. Harrassowitz.
- (1995). “Towards a reassessment of Uzbekistan Arabic”. In: *Proceedings of the 2nd international conference of l’Association Internationale pour la Dialectologie Arabe. Held at Trinity Hall in the University of Cambridge 10–14 September 1995*. Cambridge: University Publications Center.
- (2006). “Iraq”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 2. Leiden: Brill, pp. 414–424.
- Jiha, Michel (1964). *Der Arabische Dialekt von Bishmizzin*. Wiesbaden: Franz Steiner Verlag.
- Johnstone, T. M. (1975). In: *Annals of Leeds University Oriental Society* 3, pp. 89–109.
- Joseph, Brian D. (2012). “A Variationist Solution to Apparent Copying Across Related Languages”. In: *Copies versus cognates in bound morphology*. Ed. by Martine Robbeets and Lars Johanson. Leiden: Brill.
- Joseph, Brian D. and Richard D. Janda (2003). “On Language, Change and Language Change — Or, of History, Linguistics, and Historical Linguistics”. In: *The Handbook of Historical Linguistics*. Ed. by Brian D. Joseph and Richard D. Janda. Malden, MA: Blackwell.
- Al-Jundī, Ahmad ‘Āli al-Dīn (1983). *al-Lahağāt al-‘Arabiyya fī l-Turāth [Arabic dialects in history]*. al-Dār al-‘Arabiyya li-l-Kitāb.
- Kahane, Henry (1986). “A Typology of the Prestige Language”. English. In: *Language* 62.3, pp. 495–508.
- Karyolemou, Marilena (In press). “Aspects of identity in the Arab community of Cyprus”. In: *Identitat, Europa, Mediterrània Dinàmiques identitàries a la Mediterrània*. Ed. by Joan Argenter. Retrieved from <. Βαρκελόνη: Càtedra UN-ESCO de Llengües i Educació Institut d’Estudis Catalans.
- Kaye, Alan S. (1976). *Chadian and Sudanese Arabic in the Light of Comparative Arabic Dialectology*. The Hague: Mouton.
- Kennedy, Hugh (1998a). “Egypt as a province in the Islamic caliphate, 641–868”. In: *Islamic Egypt, 640–1517*. Ed. by Carl F. Petry. Cambridge: Cambridge University Press.
- (1998b). “From Antiquity to Islam in the Cities of al-Andalus and al-Mashriq”. In: *Genèse de la ville islamique en al-Andalus at au maghreb occidental*. Ed. by P. Cressier and M. Garcia-Arenal. Madrid: CSIC.
- (2001). “Map: Arabia in the Time of the Hījra (622 A.D.)” In: *An Historical Atlas of Islam*. Ed. by Hugh Kennedy. Brill Online.

- Kennedy, Hugh (2007). *The Great Arab Conquests: How the Spread of Islam Changed the World We Live In*. London: Weidenfeld and Nicolson.
- Kerr, Robert M. (Feb. 2013). *The Language of the Koran*. Retrieved 6-18-2013. Tingus Redux. URL: http://www.tingisredux.com/article/the_language_of_the_koran.html.
- Kerswill, Paul (2010). "Contact and New Varieties". In: *The Handbook of Language Contact*. Wiley-Blackwell, pp. 230–251.
- Kerswill, Paul and Ann Williams (2000). "Creating a New Town koine: Children and language change in Milton Keynes". In: *Language in Society* 29, pp. 65–115.
- Khalfaoui, Amel (2007). "A cognitive approach to analyzing demonstratives in Tunisian Arabic". Unpublished Ph.D. dissertation. University of Minnesota.
- (Feb. 2013). "A Salience-based Analysis of the Tunisian Arabic Demonstrative *hāk* as Used in Oral Narratives". Presented at the 27th Arabic Linguistics Symposium, held February 28th–March 2nd, 2013 at Indiana University in Bloomington.
- Khan, G. (1997). "The Arab Dialect of the Karaite Jews of Hīt". In: *Zeitschrift für arabische Linguistik* 34, pp. 53–102.
- King, G.R.D. (1994). "Settlement in Western and Central Arabia and the Gulf in the Sixth-Eighth Centuries A.D." In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.
- Kitchen, Andrew et al. (2009). "Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East". In: *Proceedings of the Royal Society B: Biological Sciences* 276.1668, pp. 2703–2710.
- Knauf, Ernst (2011). "Arabo-Aramaic and 'Arabiyya: From Ancient Arabic to Early Standard Arabic, 200 CE-600 CE". In: *The Qur'an in Context: Historical and Literary Investigations into the Qur'anic Milieu*. Ed. by Angelika Neuwirth, Nicolai Sinai, and Michael Marx. Leiden: Brill.
- Kremer, Ludger (1990). "Kontinuum or Bruchstelle? Zur Entwicklung der Grenzdialekte zwischen Vechtegebeit und Niederrhein". In: *Germanistische Linguistik* 101.3, pp. 85–123.
- Kristiansen, G. (2003). "How to do things with allophones: linguistic stereotypes as cognitive reference points in social cognition." In: *Cognitive Models in Language and Thought*. Ed. by R Dirven, R Frank, and M. Pütz. Berlin: Mouton De Gruyter, pp. 69–120.

- Kubiak, Wladyslaw B. (1987). *Al-Fustat: Its Foundation and Early Urban Development*. Cairo: The American University in Cairo Press.
- Labov, William (1963). "The social motivation of a sound change". In: *Word* 19, pp. 273–309.
- (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Malden, MA: Blackwell.
- (2011). *Principles of Linguistic Change, Volume 3: Change, Cognitive and Cultural Factors*. Malden, MA: Wiley-Blackwell.
- Lecker, M. (2012). "Tamīm b. Murr". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Lentin, Jérôme (2006a). "Damascus Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 546–555.
- (2006b). "Middle Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 87–96.
- LePage, Robert (1992). "'You can never tell where a word comes from': language contact in a diffuse setting". In: *Language contact: Theoretical and empirical studies*. Ed. by Ernest Håkon Jahr. Berlin: Mouton De Gruyter.
- LePage, Robert and Andrée Tabouret-Keller (1985). *Acts of Identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Llamas, Carmen (2007). "'A Place between Places': Language and Identities in a Border Town". English. In: *Language in Society* 36.4, pp. 579–604.
- (2010). "Convergence and Divergence Across a National Border". In: *Language and Identities*. Ed. by Carmen Llamas and Dominic Watt. Edinburgh: Edinburgh University Press.
- Llamas, Carmen, Dominic Watt, and Daniel Ezra Johnson (2009). "Linguistic Accommodation and the Salience of National Identity Markers in a Border Town". In: *Journal of Language and Social Psychology* 28.4, pp. 381–407.
- Luttrell, A. T. (1991). "Medieval Malta: The non-written and the written evidence". In: *Malta: A Case Study in International Cross Currents*. Malta University Publications, pp. 33–45.
- MacAdam, Henry Innes (1994). "Settlement and Settlement Patterns in Northern and Central Transjordan ca 550-ca 750". In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.

- Macdonald, M.C.A (2000). "Reflections on the linguistic map of pre-Islamic Arabia". In: *Arabian Archaeology and Epigraphy* 11, pp. 28–79.
- (2006). "Old Arabic (Epigraphic)". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Leiden: Brill.
- (2008a). "Ancient North Arabian". In: *The Ancient Languages of Syria-Palestine and Arabia*. Ed. by Roger D. Woodward. Cambridge: Cambridge University Press.
- (2008b). "The decline of the 'epigraphic habit' in late antique Arabia: some questions". In: *L'Arabie à la veille de l'Islam, bilan clinique : Actes de la table ronde*. Ed. by Jérémie Schiettecatte and Christian Robin. Orient & Méditerranée 3. Editions DE BOCCARD.
- (2009a). "Arabs, Arabias and Arabic Before Late Antiquity". In: *Topoi* 16.
- (2009b). *Literacy and Identity in Pre-Islamic Arabia*. Burlington, VT: Ashgate.
- (2010). "Ancient Arabia and the written word". In: *The Development of Arabic as a Written Language: Papers from the Special Session of the Seminar for Arabian Studies held on 24 July, 2009*. Ed. by M. C. A. MacDonald. Supplement to the proceedings of the Seminar for Arabian Studies Volume 40. Archaeopress.
- Mahdi, Qasim R. (1985). "The spoken Arabic of Baṣra, Iraq". Unpublished Ph.D. thesis. The University of Exeter.
- Marçais, Philippe (1956). *Le parler arabe de Djidjelli, Nord constantinois, Algérie*. Paris: Librairie d'Amérique et d'Orient.
- Marçais, William (1902). *Le Dialecte Arabe Parlé a Tlemcen: Grammaire, Textes et Glossaire*. Paris: Publications de L'École des Lettres D'Alger.
- (1938/1956). "Comment L'afrique Du Nord A Été Arabisée". In: *Annales de l'Institut d'études orientales d'Alger* IV/XIV. Originally two articles, pp. 171–184 published 1938, pp. 184–192 published 1956., pp. 171–192.
- Matras, Yaron (2007). "The borrowability of structural categories". In: *Grammatical Borrowing In Cross-Linguistic Perspective*. Ed. by Yaron Matras and Jeanette Sakel. Walter de Gruyter.
- (2009). *Language Contact*. Cambridge: Cambridge University Press.
- Matras, Yaron and Maryam Shabibi (2007). "Grammatical Borrowing in Khuzistani Arabic". In: *Grammatical Borrowing in Cross Linguistic Perspective*. Ed. by Yaron Matras and Jeanette Sakel. Berlin: Mouton de Gruyter.
- Mayerson, Philip (1986). "The Saracens and the Limes". In: *Bulletin of the American Schools of Oriental Research* 262, pp. 35–47.

- McWhorter, J. (2007). *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford scholarship online. Oxford University Press, USA.
- Mesthrie, Rajend (1992). *Language in indenture: a sociolinguistic history of Bhojpuri-Hindi in South Africa*. London: Routledge.
- Michalski, Marcin (2011). "The Dual Distal Demonstrative in Modern Standard Arabic: Between Norms and Non-Canonicity." In: *Lingua Posnaniensis* LIII.2, 61–73.
- Mifsud, Manwel (1995). *Loan Verbs in Maltese: A Descriptive and Comparative Study*. Leiden: E.J. Brill.
- (2006). "Maltese". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 146–159.
- Miller, Ann M. (1986). "The Origin of the Modern Arabic Sedentary Dialects: An Evaluation of Several Theories". In: *Al-Arabiyya* 19, pp. 47–74.
- Miller, Catherine (2005). "Between accomodation and resistance: Upper Egyptian migrants in Cairo". In: *Linguistics* 43.5, pp. 903–956.
- Milroy, James and Lesley Milroy (1985). "Linguistic Change, Social Networks and Speaker Innovation". In: *Journal of Linguistics* 21.2, pp. 339–384.
- Milroy, Lesley (1980). *Language and Social Networks*. Oxford: Blackwell.
- (2008). "Social Networks". In: *Handbook of Language Variation and Change*. Ed. by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes. Wiley-Blackwell.
- Mitchell, T. F. (1962). *Colloquial Arabic, the living language of Egypt*. London: English Universities Press.
- Monroe, James T. (1972). "Oral Composition in Pre-Islamic Poetry: The Problem of Authenticity". In: *Journal of Arabic Literature* 3.1, pp. 1–53.
- Morony, Michael G. (1984). *Iraq after the Muslim Conquest*. Princeton, NJ: Princeton University Press.
- (1994). "Land Use and Settlement Patterns in Late Sasanian and Early Islamic Iraq". In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.
- Moscatti, Sabatino et al. (1964). *An introduction to the Comparative Grammar of The Semitic Languages: Phonology and Morphology*. Ed. by Sabatino Moscati. Wiesbaden: Otto Harrassowitz.

- al-Mubarrad, Abī al-‘Abbās Muḥammad ibn Yazīd (1965-1968). *Kitāb al-muqtaḍab, ṣan‘at Abī al-‘Abbās Muḥammad ibn Yazīd al-Mubarrad*. Ed. by ‘Abd al-Khāliq ‘Uḍaymah. Author died. 898 CE. Cairo: Egyptian Ministry of Awqaf.
- Mufwene, Salikoko S. (2004). “Language Birth and Death”. In: *Annual Review of Anthropology* 33.1, pp. 201–222.
- Mühlhäusler, Peter (1996). *Linguistic ecology: Language change and linguistic imperialism in the Pacific region*. London: Routl.
- Naīm, Samia (2006). “Beirut Arabic”. In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 274–286.
- Nebel, A. et al. (2002). “Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa”. In: *American journal of human genetics* 70.6. Letter to the Editor, p. 1594.
- Nichols, Johanna (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna and Tandy Warnow (2008). “Tutorial on Computational Linguistic Phylogeny”. In: *Language and Linguistics Compass* 2.5, pp. 760–820.
- Niedzielski, Nancy (1999). “The Effect of Social Information on the Perception of Sociolinguistic Variables”. In: *Journal of Language and Social Psychology* 18.1, pp. 62–85.
- Noonan, Michael (2001). “The ‘Double Demonstratives’ of Chantyal”. In: *Linguistics of the Tibeto-Burman Area* 24.2, pp. 173–189.
- Noth, Albrecht (2004). “Problems of Differentiation between Muslims and Non-Muslims: Re-reading the ‘Ordinances of ‘Umar’”. In: *Muslims and Others in Early Islamic Society*. Ed. by Robert Hoyland. trans. by M. Muelhaeusler from ‘Abgrenzungsprobleme zwischen Muslimen und Nicht-Muslimen. Die ‘Bedingungen ‘Umars (aš-Šurūṭ al-umariyya)’ unter einem anderen Aspekt gelesen,’ Jerusalem Studies in Arabic and Islam 9 (1987), pp. 290-315. Ashgate Variorum.
- Ostler, Nicholas (2005). *Empires of the Word: A Language History of the World*. London: HarperCollins.
- Owens, Jonathan (1984). *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrassowitz.
- (1993). *A Grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz Verlag.
- (2003). “Arabic Dialect History and Historical Linguistic Mythology”. In: *Journal of the American Oriental Society* 123.4, pp. 715–740.

- Owens, Jonathan (2005). "Pre-diaspora Arabic: Dialects, statistics and historical reconstruction". In: *Diachronica* 22.2, pp. 271–308.
- (2006). *A Linguistic History Of Arabic*. Oxford: Oxford University Press.
- Palva, Heikki (2006a). "Dialects: Classification". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Vol. 1. Leiden: Brill, pp. 604–611.
- (2006b). "Dialects: Classification". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Vol. 1. Leiden: Brill, pp. 604–611.
- (2009). "From *qaltu* to *ḡalāt*: Diachronic Notes on Linguistic Adaptation in Muslim Baghdad Arabic". In: *Arabic Dialectology: In Honor of Clive Holes on the Occasion of his Sixtieth Birthday*. Ed. by Enam al-Wer and Rudolf de Jong. Leiden: Brill.
- Papaconstantinou, Arietta (2012). "Why did Coptic fail where Aramaic succeeded?" In: *Multilingualism in the Graeco-Roman Worlds*. Ed. by Alex Mullen and Patrick James. Cambridge University Press.
- Pat-El, Na'ama (2009). "The Development of the Semitic Definite Article: A Syntactic Approach". In: *Journal of Semitic Studies* LIV.1, pp. 19–50.
- Pellat, Ch. (2012). "al-Ma'ḳil". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Penny, Ralph J. (2004). *Variation and Change in Spanish*. Cambridge: Cambridge University Press.
- Pereira, Christophe (2006a). "Libya". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 3. Leiden: Brill, pp. 52–58.
- (2006b). "Tripoli Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 548–556.
- Pommerol, Patrice Jullien de (2006). "Chad Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 360–368.
- Preston, Dennis R. (2013). "The influence of regard on language variation and change". In: *Journal of Pragmatics* 52, pp. 93–104.
- Procházka, Stephan (2004). "Unmarked Feminine Nouns in Modern Arabic Dialects". In: *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of his Sixtieth Birthday*. Ed. by Kees Versteegh and Martine Haak Rudolf de Jong. Brill.

- Procházka, Stephan (2006). "Cilician Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 388–397.
- Procházka Jr., Theodore (1988). "Gleanings from southwestern Saudi Arabia". In: *Zeitschrift für Arabische Linguistik* 19.1, pp. 44–49.
- Procházka Jr., Theodore. (1988). *Saudi Arabian Dialects*. London: Kegan Paul International.
- Pulgram, Ernst (1959). "Proto-Indo-European Reality and Reconstruction". English. In: *Language* 35.3, pp. 421–426.
- (1961). "The nature and use of proto-languages". In: *Lingua* 10, pp. 18–37.
- Rabin, Chaim (1951). *Ancient West-Arabian*. Tylors Foreign Press.
- Al-Rawas, Isam (2000). *Oman in Early Islamic History*. Ithaca Press.
- Razzūq, Muḥammad (1989). *al-Andalusīyūn wa-hijrātuhum ilā al-Maghrib : khilāla al-qarnayn 16-17*. Casablanca: Afriqiyā al-Sharq.
- Reichmuth, Stefan (1983). *Der arabische Dialekt der Šukriyya im Ostsudan*. New York: Georg Olms Verlag.
- Retsö, Jan (2002). *The Arabs in Antiquity: Their History from the Assyrians to the Umayyads*. Routledge.
- (2003). *The Arabs in antiquity: their history from the Assyrians to the Umayyads*. London: RoutledgeCurzon.
- (2006). "Arab". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Leiden: Brill.
- Reynolds, Dwight F. (1989). "Sirat Banī Hilāl: Introduction and Notes to an Arab Oral Epic Tradition". In: *Oral Tradition* 4.1-2, pp. 80–100.
- Richter, Tonio Sebastian (2009). "Greek, Coptic and the 'language of the Hijra': the rise and decline of the Coptic language in late antique and medieval Egypt". In: *From Hellenism to Islam: Cultural and Linguistic Change in the Roman Near East*. Ed. by Hannah Cotton et al. Cambridge: Cambridge University Press.
- Riddle, Elizabeth M. (2008). "Complexity in isolating languages: Lexical elaboration versus grammatical economy". In: *Language Complexity: Typology, Contact, Change*. Ed. by Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson. John Benjamins Publishing Company.
- Robin, C. (1991). "La pénétration des Arabes nomades au Yémen". In: *Revue du Monde Musulman et de la Méditerranée* 61.1, pp. 71–88.

- Robin, Christian Julien (2006). "Ḥimyaritic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 2. Leiden: Brill, pp. 256–261.
- Rosenhouse, J (1984). *The Bedouin Arabic dialects: General problems and a close analysis of North Israel Bedouin dialects*. O. Harrassowitz.
- Ross, Malcolm (1997). "Social Networks and Kinds of Speech-Community Event". In: *Archaeology and Language I: Theoretical and Methodological Orientations*. London: Routledge.
- (2005). "Diagnosing prehistoric language contact". In: *Motives for language change*. Ed. by Raymond Hickey. Cambridge: Cambridge University Press.
- (2007). "Calquing and Metatypy". In: *Journal of Language Contact* THEMA 1, pp. 116–143.
- Roth, Arlette (1979). *Esquisse grammaticale du parler arabe d'Abbéché, Tchad*. Paris: Librairie Orientaliste Paul Geuthner.
- Rubin, Aaron D. (2005). *Studies in Semitic Grammaticalization*. Eisenbrauns.
- al-Sāmarrā'i, Ibrāhīm (1994). *Fī al-Lahağāt al-‘Arabiyya al-Qadīma [On the Old Arabic dialects]*. Beirut: Dār al-Ḥadāṭa.
- Sajīnī, Fāyizah Muḥammad Ṣāliḥ Amīn. (2007). *Ghazw Banī Hilāl wa-Banī Sulaym lil-Maghrib [Bani Hilal and Bani Sulaym's Conquest of the Maghreb]*. Cairo: Maktabat al-Nahḍah al-Miṣriyah.
- Salonen, Erkki (1980). "On the Arabic Dialect Spoken in Širqāt (Assur)". In: *Suomalaisen tiedeakatemia toimituksia annales academiae scientiarum fennicae* 212.
- Al-Saqqaf, Abdullah Hassan (2006). "Wādī Ḥaḍramawt Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 687–699.
- Scagliarini, Florella (2007). "The word ṣlm/ṣnm and some words for "statue, idol" in Arabian and other Semitic languages". English. In: *Proceedings of the Seminar for Arabian Studies* 37, pp. 253–262.
- Schabert, Peter (1976). *Laut- und Formenlehre des Maltesischen anhand zweier Mundarten*. Verlag Palm & Enke Erlangen.
- Schiettecatte, J. (2010). "Why did the cities of the Jawf valley collapse? An archaeogeographical approach". In: *Regards croisés sur l'étude archéologique des paysages anciens. Nouvelles recherches dans le Bassin méditerranéen, en Asie Centrale et au Proche et au Moyen-Orient*, pp. 149–161.
- Schleife, J. (2012). "Hilāl". In: *Encyclopedia of Islam, Second Edition*. Brill Online.

- Schumann, John H. (1978). "The acculturation model for second language acquisition". In: *Second language acquisition and foreign language teaching*. Ed. by R. Andersen. Arlington, VA: Center for Applied Linguistics.
- (1986). "Research on the Acculturation Model for Second Language Acquisition". In: *Journal of Multilingual and Multicultural Development* 7.5, pp. 379–392.
- Sebag, P. (2012). "Tūnis". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Seeger, Ulrich (1996). *Der arabische Dialekt von il-Xalil (Hebron)*. Possibly unpublished manuscript. Karlsruhe.
- Shachmon, O. (2013). "hūta, hīta and other extended pronouns in Palestinian Arabic". In: *Zeitschrift für Arabische Linguistik* 57, pp. 70–82.
- Shahīd, Irfan (1984). *Byzantium and the Arabs in the Fourth Century*. Dumbarton Oaks Research Library and Collection.
- (1989). *Byzantium and the Arabs in the Fifth Century*. Dumbarton Oaks Research Library and Collection.
- (1995). *Byzantium and the Arabs in the sixth century*. Washington, D.C.: Dumbarton Oaks Research Library and Collection.
- (2012a). "Ḡhassān". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- (2012b). "Salīḥ". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- (2012c). "Salīḥ". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- al-Sharkawi, Muhammad (2010). *The Ecology of Arabic: A study of Arabicization*. Brill.
- Shaw, Brent D. (1982). "'Eaters of Flesh, Drinkers of Milk': The Ancient Mediterranean Ideology of the Pastoral Nomad". In: *Ancient Societies* 13, pp. 5–31.
- El-Sibai, Mirvat et al. (2009). "Geographical Structure of the Y-chromosomal Genetic Landscape of the Levant: A coastal-inland contrast". In: *Annals of Human Genetics* 73.6, pp. 568–581.
- Siegel, Jeff (1985). "Koinés and koinéization". In: *Language in Society* 14.3, pp. 357–78.
- (1987). *Language Contact in a Plantation Environment*. Cambridge: Cambridge University Press.
- Sibawayhi (1999). *al-Kitāb*. Ed. by Imīl Badī' Ya'qūb. Beirut: Dār al-Kutub al-'Ilmiyya.
- Silva-Corvalán, Carmen (2008). "The limits of convergence in language contact". In: *Journal of Language Contact* THEMA 2, pp. 213–224.
- Singer, Hans-Rudolf (1984). *Grammatik der arabischen Mundart der Medina von Tunis*. Berlin: Walter de Gruyter.

- Stenström, A.B., G. Andersen, and K.I. Hasund (2002). *Trends in Teenage Talk: Corpus Compilation, Analysis and Findings*. Amsterdam: John Benjamins Publishing Company.
- Taeldeman, Johan (1995). "The Influence of Urban Centres on the Spatial Diffusion of Dialect Phenomena". In: *Dialect Change: Convergence and Divergence in European Languages*. Ed. by Peter Auer, Frans Hinskens, and Paul Kerswill. Cambridge: Cambridge University Press.
- Ṭāhā, 'Abd al-Wāḥid Dhannūn (1989). *The Muslim conquest and settlement of North Africa and Spain*. New York: Routledge.
- Taine-Cheikh, Catherine (2006). "Ḥassāniyya Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 687–699.
- Al-Tajir, Mahdi Abdalla (1982). *Language and Linguistic Origins in Bahrain*. London: Kegan Paul International.
- Talay, Shabo (1999). *Der arabische Dialekt der Khawētna*. Harrassowitz Verlag.
- Talmoudi, Fathi (1980). *The Arabic Dialect of Sūsa (Tunisia)*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.
- Tannous, Jack Boulos Victor (2010). "Syria Between Byzantium and Islam: Making Incommensurables Speak". Unpublished Ph.D. Dissertation. Princeton University.
- Testen, David Douglas (1998). *Parallels in Semitic Linguistics: The Development of Arabic La-And Related Semitic Particles*. Studies in Semitic Languages and Linguistics, 26. Leiden: Brill Academic Publishers.
- Thomason, Sarah Grey (2007). "Language Contact and Deliberate Change". In: *Journal of Language Contact* 1.1, pp. 41–62.
- Thomason, Sarah Grey and Terrence Kaufman (1988). *Language Contact, Creolization, and Genetic Linguistics*. Los Angeles: University of California Press.
- Toulmin, Matthew (2009). *From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University's Center for Research on Language Change.
- (2012). "Historical Sociolinguistic Reconstruction Beyond Europe: Case Studies from South Asia and Fiji". In: *The Handbook of Historical Sociolinguistics*. Ed. by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre. Wiley-Blackwell.
- Trudgill, Peter (1986). *Dialects in Contact*. Oxford: Basil Blackwell.

- Trudgill, Peter (2004). *New Dialect Formation: The Inevitability of Colonial Englishes*. Oxford: Oxford University Press.
- (2008). "Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation". In: *Language in Society* 37.02, pp. 241–254.
- Trudgill, Peter and Paul Kerswill (2005). "The Birth of New Dialects". In: *Dialect Change: Convergence and Divergence in European Languages*. Ed. by Peter Auer, Frans Hinskens, and Paul Kerswill. Cambridge: Cambridge University Press.
- Tsafir, Yoram and Gideon Foerster (1992). "The Dating of the 'Earthquake of the Sabbatical Year' of 749 C. E. in Palestine". In: *Bulletin of the School of Oriental and African Studies* 55.2, pp. 231–235.
- (1994). "From Scythopolis to Baysan - Changing Concepts of Urbanism". In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.
- Vanhove, Martine (1995). "Notes On The Arabic Dialectal Area Of Yāfi' (Yemen)". In: *Proceedings of the Seminar for Arabian Studies Proceedings of the Seminar for Arabian Studies Vol. 25, Papers from at the twenty-eighth meeting of the Seminar For Arabian Studies held at Oxford, 21-23 July 1994 (1995)*, pp. 141–152.
- (2004). "Deixis Et Focalisation: La Particule TA En Arabe De Yafi' (yémen)". In: *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*. Ed. by Martine Haak, Rudolf Erik De Jong, and Kees Versteegh. Leiden: Brill.
- Versteegh, Kees (1984). *Pidginization and Creolization: The Case of Arabic*. Amsterdam: Benjamins.
- ed. (2005 – 2009). *Encyclopedia of Arabic Language And Linguistics*. Associate Editors: Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborski. Leiden, the Netherlands: Brill Academic Publishers.
- (2006). "Poetic Koine". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Web. 20 Dec. 2011: Gale Virtual Reference Library.
- Waines, David (1977). "The Third Century Internal Crisis of the Abbasids". In: *Journal of the Economic and Social History of the Orient* 20.3, pp. 282–306.
- Wasserstein, David (1998). "Language and Religion: The language situation in al-Andalus". In: *The Formation of al-Andalus, Part 2: Language, Religion, Cul-*

- ture and the Sciences. Ed. by Maribel Fierro and Julio Samsó. Ashgate Variorum.
- Watson, Janet C. E. et al. (2006a). "The language of Jabal Rāziḥ: Arabic or something else?" In: *Proceedings of the Seminar for Arabian Studies Vol. 36, Papers from the thirty-ninth meeting of the Seminar for Arabian Studies held in London, 21-23 July 2005 (2006)*. Oxford: Archaeopress, pp. 35–41.
- (2006b). "Two Texts from Jabal Rāziḥ, North-west Yemen". In: *Current Issues in the Analysis of Semitic Grammar and Lexicon II: Oslo-Göteborg Cooperation 4th-5th November 2005*. Ed. by Lutz Edzard and Jan Retsö. Wiesbaden: Harrassowitz Verlag.
- Watt, Dominic, Carmen Llamas, and Daniel Ezra Johnson (2010). "Levels of Linguistic Accommodation across a National Border". In: *Journal of English Linguistics* 38.3, pp. 270–289.
- Weijnen, A. (1977). *The Value of the Map Configuration*. Special issue of "Mededelingen van de Nijmeegse Centrale voor Dialect- en Naamkunde". Nijmegen: Katholieke Universiteit Nijmegen.
- Al-Wer, Enam (2004). "Variability Reproduced: A Variationist View of the [d]/[d̥] opposition in modern Arabic dialects". In: *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of his Sixtieth Birthday*. Brill.
- (2007). "The Formation of the Dialect of Amman: From Chaos to Order". In: *Arabic in the City*. Ed. by Catherine Miller et al. Routledge.
- Whitcomb, Donald (1994). "The Misr of Ayla: Settlement at al-ʿAqaba in the Early Islamic Period". In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.
- Wikle, Thomas and Guy Bailey (1996). "The Spatial Diffusion of Linguistic Features in Oklahoma". In: *Proceedings of the Oklahoma Academy of Science* 77.1, pp. 1–15.
- Wikle, Thomas, Guy Bailey, et al. (1993). "Some Patterns of Linguistic Diffusion". In: *Language Variation and Change* 3.3, pp. 241–64.
- Winford, Donald (2003). *An Introduction to Contact Linguistics*. Oxford: Blackwell.
- Woidich, Manfred (2006a). "Bʿeri Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 299–308.
- (2006b). *Das Kairenisch-Arabische. Eine Grammatik*. Wiesbaden: Harrassowitz Verlag.

- Wolff, Hans (1959). "Intelligibility and Inter-Ethnic Attitudes". In: *Anthropological Linguistics* 1.3, pp. 34–41.
- Wright, W. (1896-1898). *Arabic Grammar*. 2005 single volume. In two parts, Part I published 1896, Part II 1898. Mineola, NY: Dover Publications, Inc.
- Yoda, Sumikazu (2005). *The Arabic Dialect of the Jews of Tripoli (Libya)*. Wiesbaden: Harrassowitz Verlag.
- Yver, G. (2012). "Djидjelli". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Yver, G. and J. Sari (2012). "Sharshal". In: *Encyclopaedia of Islam, Second Edition*. Brill Online.
- Zack, Elisabeth (2009). "Egyptian Arabic in the Seventeenth Century: A study and edition of Yūsuf al-Mağribī's Daf' al-iṣr 'an kalām ahl Miṣr". PhD thesis. University of Amsterdam.
- Zack, Elisabeth and Arie Schippers, eds. (2012). *Middle Arabic and Mixed Arabic: Diachrony and Synchrony*. Leiden: Brill Academic Publishers.
- Zelinsky, Wilbur (1992). *The Cultural Geography of the United States: A Revised Edition*. Englewood Cliffs, NJ: Prentice Hall.
- Zewi, Tamar (2006). "Diminutives". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 637–640.
- Zeyadeh, Ali (1994). "Settlement Patterns, An Archaeological Perspective. Case Studies from Northern Palestine and Jordan". In: *The Byzantine and Early Islamic Near East II: Land Use and Settlement Patterns*. Ed. by G.R.D. King and Averil Cameron. Princeton, New Jersey: The Darwin Press, Inc.
- Zimmermann, Gerit (2006). "Uzbekistan Arabic". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 4. Leiden: Brill, pp. 612–623.
- Zwartjes, Otto (2006). "Andalus". In: *Encyclopedia of Arabic Language and Linguistics*. Ed. by Kees Versteegh. Online edition. Vol. 1. Leiden: Brill, pp. 92–101.