Pre-diaspora Arabic
Dialects, statistics and historical reconstruction *

Jonathan Owens
University of Maryland

Arabic dialects, the native spoken Arabic of about 250 million people, are spread over an immense, contiguous geographical area from Iran to Lake Chad, from Morocco to Yemen. Corresponding to this geographical spread is considerable linguistic diversity. An explanation for this diversity has proved elusive. The existence of variants found either in the modern dialects or in the Classical literature (or both), which are not self-evidently derivable from a normalized Classical Arabic (largely standardized by the ninth century), argues for a more diverse set of inputs into the Arabic which spread outwards from the Arabian Peninsula beginning in the seventh century. I elucidate this problem by comparing four varieties of Arabic located in widely separated areas and settled at different times. To account for the internal diversity of the areas compared, a dataset is established with 49 phonological and morphological features, which, using simple statistical procedures, permits a normalized comparison of the varieties. From this set of variables, two specific linguistic features are discussed in detail and reconstructions proposed, which place their origins in a pre-diaspora variety. I conclude that the Arabic which preceded the Arabic diaspora of the seventh century was considerably more diverse than interpretations of the history of Arabic traditionally allow for.

Keywords: History of Arabic, Arabic dialectology, statistics and comparative grammar

* My thanks go to three readers for their detailed comments on an earlier draft of this paper.
1. Introduction

In this paper I treat a problem in historical linguistics using a set of statistical data. The historical question relates to the nature of early Arabic. In particular I am interested in the form of Arabic spoken in the early Islamic period, the period immediately preceding the great Arab diaspora which began with the expansion if Islam, what I term 'pre-diaspora Arabic'. I use the date 640 c.e. as a convenient point of orientation, cognizant that expansion had begun before this time in the pre-Islamic era for the lands immediately adjacent to the Arabian Peninsula (see Retsö 2003 for a comprehensive summary).

Direct, contemporary evidence pertaining to this variety is rare and incomplete. Assuming standard interpretations of the history of the Quranic text, the so-called muṣlaʃ, there was a written document by 652 at the latest (the Uthmanic codex). However, this consisted of only the barest consonantal skeleton with no short vowels marked, and even no diacritical points to distinguish consonants. Initial t and n, for instance, are formally indistinguishable. Moreover, full Qurans from this period do not exist, only early excerpts (Puin 1996). It is not until the early eighth century that complete Quranic renditions, with all consonants and vowels spelled out, become available in the tradition of Quranic reading (qiraa'aat). Even these, however, are not fully systematized and written down until the 10th century by Ibn Mujahid (d. 936/324), and so it is not until this date that fixed, complete versions of the Quran are accessible to us, representing all vowels and unambiguously identifying all consonants. Thus, as far as early written sources go, the earliest versions of the Quran, even if they trace a chain (nisnaad) back to the early Islamic period, do not become linguistically explicit until the tenth century.

Moreover, this Quranic tradition is, excepting the relatively small amount of epigraphic material, the only one available for the earliest Arabic. Even under the most favorable assumption that it provides direct insights into the Arabic of 640, it cannot be claimed to be representative of all varieties of Arabic of the early period. This point becomes apparent as early as the explicitly defined reading tradition just mentioned, where it is clear that different traditions draw on different varieties of Arabic. An extended example of this is so-called 'imala', discussed in Section 4.2. There was not one variety of early Arabic, but many.

In the tradition of Arabic linguistics, these varieties are known as early or 'Old' or 'ancient' Arabic dialects (e.g. Brockelmann 1908; Rabin 1951). This designation is not very helpful without linguistic content, but, to my knowledge, no one has yet provided that. Rabin (1951) comes closest, though he lim-
its himself to what he calls "west Arabian", and his work, though valuable, does not exhaust even the written material from the classical era (e.g. he does not adduce evidence from the qiraa’i‘at), nor can the features which he describes be said to constitute a comprehensive dialectology, in a modern sense. Often, as with Fischer (1995), an Old Arabic dialect is simply shorthand for whatever is not Classical Arabic.\footnote{I use the term ‘Classical Arabic’ as a cover term for the Arabic of the Quran, that described by the Arabic grammarians and the variety as it came to be used in the course of the ninth century. It therefore includes varieties which are sometimes distinguished as Quranic Arabic, pre-classical Arabic, and Classical Arabic.}

In fact, the concept of ‘dialect’, whether old or contemporary, is a very important one for understanding Arabic linguistic history, for a simple reason. A great deal of the information we have about Arabic is popularly attributed, rightly or wrongly, to the category of ‘dialect’. Not to incorporate this information systematically into an account of Arabic linguistic history is to write half a history of the Arabic language.

Because it is so important, let us pause here to consider four aspects of the term ‘Arabic dialect’. First, there is the issue of what constitutes information on an Arabic dialect. In the contemporary era, I believe it is fairly uncontroversial that dialect simply means the L1 of native Arabic speakers. These native varieties may be differentiated by classic dialectological methods, using bundles of isoglosses to define dialect areas. As may be gleaned from the previous discussion, it is far more difficult to agree on what constitutes a ‘dialect’ in the classical era. A basic problem is that from the perspective of the on-site observers in the late eighth and early ninth centuries, linguists like Sibawaih (d. 793/177) and Farra’ (d. 822/207), the distinction between dialect and Classical Arabic did not exist as it came to be understood by (relatively) modern observers such as Brockelmann or Fischer. Thus, a feature which by modern standards such as the 2r.sg object suffix -\textit{wi}, as in \textit{inma-fi “that-you.\textit{f}”} is clearly ‘dialectal’ was simply treated by Sibawaih as an explicable morphophonological variant of -\textit{ki} (II:322). This point is a large issue which can only be adumbrated here. Without arguing the point in detail, my position is that it is necessary methodologically to distinguish between linguistic and sociolinguistic aspects of Arabic for purposes of interpreting Arabic linguistic history. I put all linguistic material, whether that which became canonized in the classical language (the \textit{fush\textst{aa}f}),
or that which may popularly be understood as dialectal, potentially on a par for purposes of reconstructing the language history (see Owens 1998b). Sociolinguistically, not all varieties are equal, as many recent studies have shown. Sociolinguistic prestige should not, however, imply precedence in interpreting language history. As a terminological point, for reasons given in the next paragraph, I use the term ‘varieties’ of Old Arabic, rather than ‘dialect’ of Old Arabic when speaking of variants attested in old sources.

Second, it is not possible to equate material from Old Arabic with modern material for basic descriptive reasons. As far as old varieties go, while tantalizing bits and pieces of odd material from a variety of sources exist (e.g. in Sibawaih, the Quranic reading tradition, epigraphic inscriptions), there is nowhere near enough to construct an old dialectology in the sense of having a relatively complete phonological and morphological account of discrete dialects. This is one reason I use the term ‘varieties’ of Old Arabic. By contrast, fortunately, a relative surfeit of information exists on modern dialects. This has important methodological consequences in the present study. I begin the comparative study with contemporary dialects and work my way backwards because systematic sets of data can be compared across all domains of grammar.

Third, the term ‘dialect’ in the Arabic tradition of Arabic countries automatically carries a pejorative connotation due to the modern opposition between ‘dialect’ vs. ‘Standard Arabic’. Only the latter is the ‘real’ Arabic so far as official recognition in Arabic countries goes, and similarly to speak of Old Arabic ‘dialects’ is to suggest the same secondary status of a variety opposed to the authentic ‘Classical Arabic’. That is, ‘dialect’ in these terms is not so much a spatial linguistic designation as a socio-political one, and a negative one at that. This aspect of the term ‘Arabic dialect’ falls outside the scope of the present paper.

Fourth and finally, dialectology itself as a linguistic tradition can be viewed as a counterfoil to historical linguistics. It requires a synchronic spatial or socio-spatial starting point. Continuing this point, wedding dialectology and historical linguistics is of course possible, though a problem intrudes that may be termed ‘reification’. Tracing the development of a dialect, it might be assumed that a dialect is a complete, discrete entity, comparable say to a building, which

2. This caveat is important. A demonstrable innovation found in a modern dialect cannot be used to reconstruct Proto-Arabic.
remains relatively unchanged through time. Under this assumption there is a temptation to start with whatever set of features one has used to define the dialect in question, and to assume that the same set of features will cohere through time, each changing in consonance with the others.

This may not be the case, however. Indeed, from a historical perspective one has to begin with the assumption that each feature has its own history: lexis changes at a different rate from phonology, verbal morphology differently from nominal, and so on. The recognition of this is what lies behind Thomason & Kaufman’s (1988) attempt at typologies of potential rates of change in different components of grammar.

The fact that dialects do not change in a coherent, uniform fashion leads to a methodological and practical quandary. To do justice to the historical linguistic reality, one needs first to trace the development of each feature which may potentially be included in a given dialect: in a given period of time, some features will change a little, others a lot, others not at all. It is, however, not practically possible to reconstruct the large numbers of individual features which are customarily used in defining a dialect, unless one has a great deal of time at one’s disposal.

Fortunately, there are two factors which allow one to circumvent these practical constraints to a degree. The first is the advent of the computer, which allows the average linguist to create relatively large data banks with which complex sets of data can be organized manageably and quickly. The second is the reality of contemporary dialects for certain languages, Arabic in particular, and the historical inferences which may be drawn from them. Arabic is spoken as a native language by about 250 million speakers in an unbroken area stretching from Lake Chad to northern Iraq, from Mauritania to Yemen. Sprachinseln outside this area also are found, and variation across these dialects is remarkable. The diaspora to these regions began in earnest with the first Islamic conquests, around 640. Moreover, once populations had moved in opposite directions, say into northern Iraq and into southern Egypt or into Chad, contact between them was largely or wholly lost. These two factors can be used in historical interpretation.

Turning then to the present data, a relatively large number of linguistic features can be coded statistically. I use 49 features, 25 phonological and 24 morphological. These features are chosen from dialects at opposing ends of the Arabic-speaking world, on the premise that retentions in particular will be due to a common inheritance dating back to the original diaspora. Since this
diaspora dates from about 640 c.e., indirect insights into the Arabic of this early era can be achieved. In a nutshell, this is the program pursued in this paper.

Of course, as one reader emphasizes, any given shared form could arise via independent innovation. Ultimately, it has to be decided on a case by case basis what is common retention versus independent innovation. Nonetheless, the assumption followed here, appealing to Occam’s Razor, is that the unmarked case in the presence of common forms is to assume retention, and that the onus lies with showing why this does not apply. I should note that this is the simplest assumption, in particular when confronted by identical forms separated by large spatial and temporal distances, such as those under discussion in this paper.

The methodology used here has a number of ramifications, however, two of which can be made explicit in order to give a more comprehensive account of the issues involved. First, the comparison between two dialect regions will throw into relief what is understood by the very term ‘Arabic dialect’. I have selected two widely separated dialect areas for detailed scrutiny, Mesopotamia and the western Sudanic area of Africa. As noted, the populations of these two regions took separate paths of migration and would have remained out of contact with each other for about 1,200 years (about 800–2000 c.e.). By way of orientation for the non-Arabicist, this is roughly comparable to the separation between Old and contemporary English. It will become apparent, however, that the two areas differ rather dramatically in terms of their dialect coherency, at least as measured by the features chosen here, a difference which itself calls for historical reflection. Second, it will be suggested that the large-scale statistical comparison does at a certain point have to give way to positing individual linguistic features as a way of reconstructing pre-diaspora Arabic. I single out two features for greater scrutiny in this respect, complementing the statistical approach with a close reading of the philological evidence.

Initially, therefore, two separate issues inform this paper. One is dialectological, the other historical and comparative. It may in fact appear that I have contradicted myself, above arguing that the concept of ‘dialect’ conceived of as a coherent unit is antithetical to historical linguistic interpretation, but here suggesting that dialectal units are legitimate elements in historical linguistics. This apparent contradiction, however, will disappear once the statistical treatment gives the dialectal units internally differentiated character. Ultimately it will turn out that contemporary dialectology converges with reconstruction to complement the historical interpretation of pre-diaspora Arabic.
The paper is divided into six parts. §2 introduces the data and dialects used. §3 presents the basic statistics and initial interpretive results. Here it will be seen that the statistics help orientate a historical interpretation, but that a more precise interpretation requires the introduction of further types of data. This is done in §4 where two specific features are discussed in greater detail. In §5 general conclusions relating to the interpretation of Arabic linguistic history are drawn on the basis of the overall data. In an appendix (§6), a list of the variables is given and the coding of each is briefly illustrated. [In addition, the many methodological problems involved in the codification and interpretation of individual features are discussed systematically in a file available on the internet at: diachronica.org/owens. Furthermore, the complete SPSS data used in the codification can also be downloaded from there.]

2. Dialects, procedure initial results

The realization of 49 linguistic variables, all from phonology and morphology, have been compared between two dialect areas, Mesopotamia and the Western Sudanic Area (WSA), in addition to Uzbekistan and Shukriyya Arabic spoken in the eastern Sudan along the Atbara River as controls.

2.1 Linguistic variables

The variables are all basic elements in the phonology and (morpho)phonology of the dialects, the realization of $t$ or $q$, the realization (or lack thereof) of short vowels in open syllables, the realization of the 1sg person suffix in perfect verbs, the realization of the (m)pl suffix in the imperfect verb, and so on. Even the most rudimentary dialect grammar describes these elements. Their centrality ensures that they may be used in a comparison of widely separated dialects. Furthermore features are chosen which are indicative of differentiation, both between Mesopotamia and WSA and within each area. It is a reality of both dialectology and comparative linguistics that only different elements are interesting for drawing inferences about dialect classification or language change. Elements which are identically maintained, for instance the 1pl perfect verb marker - naughty, are interesting for neither. The features used are listed in §6.

There are many methodological and theoretical issues connected with the choice of these variables, and their individual realizations in the different dialects. Throughout, various of these problems are addressed, although detailed
Turning to the choice of dialect areas, four considerations are in play. First, as noted above, I compare two modern dialect areas which were settled in the course of the Arabic diaspora accompanying the spread of Islam. Second, the areas are widely enough dispersed that it is unlikely that there has been significant movement between these regions after the initial migrations out of Arabia. In general, common features should be due to common inheritance from the era of original expansion, or before. Third, areas are compared which are very roughly comparable in size. These are Mesopotamia and the Western Sudanic Area. I limit the Mesopotamian area here to Iraq and the bordering area of southern Turkey (Anatolia). Iraq has an area of 169,235 square miles, while the Arabic-speaking areas of Anatolia cover only a small area. The native Arabic-speaking population in these two areas is about 11,500,000 in Iraq and 400,000 in Turkey (Ethnologue 2003). By the WSA, I understand the western extension of the Arabic-speaking belt beginning in Kordofan and stretching into Nigeria (Owens 1993b). The total area of Chad is 499,755 square miles; Arabic is only spoken as a native language in approximately the southern third of the country. De Pommerol (1997:9) estimates that only 10% of the population are native Arabic speakers, which would give a population of about 700,000 native speakers. It is further represented as a native language in the small finger of Cameroon separating Chad and Nigeria, and in Nigeria in about half of Bornu (69,436 square miles). Estimates for the number of native Arabic speakers in these areas vary wildly, though the total is probably not more than 500,000 speakers. In all, very approximately, the area where Arabic is spoken as a native language in WSA is perhaps 250,000 square miles and the total number of speakers just over 1,000,000.

Fourth, a minimal constraint on the choice of dialects was the availability of material covering all of the linguistic variables. Beyond this, the comparison was limited to the same number of dialect samples in each area. To overweight one area or the other would vitiate some of the arguments advanced in
§3. Practically, these two considerations mean that relative to the number of descriptions available, the Mesopotamian area is underrepresented while the WSA is overrepresented.

I have used nine sample points for Mesopotamia (§2.2.1) and nine for WSA (§2.2.2). This allows for direct statistical comparability. Alternatively one could attempt to construct proportionally representative samples based on any number of parameters, such as geographical size of the two regions or total population. These would constitute alternative statistical approaches, which I will not attempt to develop here.

I have strategically added two further dialects, Uzbekistan Arabic and Shukriyya Arabic (eastern Sudan on the Atbara River). These function as controls, both in terms of defining dialect areas, and in terms of a historical interpretation of the data. For historical linguistics, Uzbekistan Arabic, as will emerge, is of particular importance, since it became cut off from the rest of the Arabic-speaking world at a very early date.

2.2.1 Mesopotamian area

The Mesopotamian area was defined as a dialect area by Haim Blanc (1964). He distinguished two large dialect groupings, dubbing them the qiltu group (here qiltu) and the gilit (here gilit). Qiltu/gilit means “I said”, and represents a form in which the characteristic differences between the two, /q/ vs. /g/ and -tu vs. -t for the 1sg perfect verb suffix, are found in a short, common word.

The Mesopotamian area, thanks in large part to the work of Otto Jastrow and his associates, is one of the better-described Arabic dialect areas, at least as far as the qiltu dialects go. A recent work (Talay 1999: 15) divides the qiltu dialects into four groups, the Anatolian, Kurdistan, Tigris and Euphrates groups. These are essentially geographically-based, but serve for present purposes as an orientation. The current sample has three members of the Anatolian group, Daragözü (= Kozluk group, Jastrow 1973), Siirt (Jastrow 1978) and Mardin (Sasse 1971; Jastrow 1978). The Anatolian group is itself quite differentiated dialectically (Jastrow 1978: Ch. 1), and the three dialects used here are all members of different, parallel subgroups. The sample has two members of the Euphrates group (Hiit and Khaweeneta, Khan 1997; Talay 1999, respectively) and two of the Tigris group (Christian = CB and Jewish Baghdad = JB; Blanc 1964; Abu Haidar 1991). The gilit dialects are underrepresented, both in this sample and in the literature in general. I have used the Muslim dialect of Baghdad (= MB, Blanc 1964; Malaika 1959), and the southern Mesopotamian dialect as described especially by Ingham (1976, 1982, see Map 1).
Blanc's summary of the introduction of these two dialects into Iraq (1964:169) is still relevant today. The ancestors of *qiltu* speakers constituted the first major wave of Arab expansion into Mesopotamia, beginning in the seventh century. Beginning in the tenth century, in the face of a breakdown of centralized government control, Iraq began increasingly to be settled by nomadic groups, who are associated with the *gilit* dialect. This historical sequence of migration explains the current distribution of dialects in contemporary Iraq: a later spread of *gilit*-speakers (1) populated previously uninhabited rural areas, (2) probably assimilated many *qiltu* dialects, and (3) fragmented the *qiltu*-speaking areas. *Qiltu*-speaking islands in northern Iraq and Anatolia, usually urban-based, would have been left unaffected by this later migration, though it is clear that a *gilit*-variety has become established as an Iraqi lingua franca.
which in Iraq at least continues to spread into qiltu dialects. I should note that while in Iraq the qiltu dialects may be regarded as the older of the two, this does not imply that qiltu dialects are, in terms of the global history of Arabic, older than the gilit. They are simply older in this area.

2.2.2 Western Sudanic Arabic

The Western Sudanic area is much more sparsely described, and some of the literature available is problematic in its interpretation. Two works of high quality are Roth (1979) for Abbeche Arabic and Zeltner & Tourneux (1986) for a dialect spoken either in far western Chad or in Cameroon. The latter authors are not explicit on the point, though they note that the phonology of the dialect they describe (identified by tribe, wulaad ‘eeli, not place) is identical to that of Zeltner & Fournier (1971). This was a description of a Cameroonian Arabic dialect. Beyond these two, I have relied on my own data. This consists of a wide range of samples from Nigerian Arabic, as well as specifically elicited information from speakers from Chad. For Nigeria I use exclusively samples from villages. The largest concentration of Arabic speakers in Nigeria is found in Maiduguri, a city of 500,000 inhabitants, with perhaps 50,000 native Arabic speakers. They come from different dialectal backgrounds, however, so it is not possible to speak of a characteristic Arabic dialect of Maiduguri (see Owens 1998a for the sociolinguistic basis of this situation). From Nigeria I use the village of Kirenawa, located on the southwest side of Lake Chad. A second Nigerian village is Mada. While this is less than 100 kilometers away, it lies in what I term the northern Bagirmi dialect area, linguistically quite distinct in certain ways from Kirenawa. The final village is Aajiri, near the town of Banki which lies at the border between Nigeria and Cameroon. It is taken as a representative of the southern Bagirmi area. To these five samples I added four from Chad via elicitation. Two of these were collected at the border town of Banki, which has a large cattle market every Wednesday and Thursday. Most of the cattle sold there are driven across Chad, through Cameroon to Banki over a period of 1–3 months, and sold to cattle dealers who transport the cattle via truck to southern Nigeria. Speakers of various Chadian dialects can therefore be found there on any market day, speakers whose residence is Chad. The speakers presented here are from Atia (Atia I) and from a nomad born south of Umm Hajar who lives nomadically between Umm Hajar and Am Timan. Two other interviews were conducted in Maiduguri, again from speakers who had recently arrived from Chad. One of these is from Amm Timan, the other a nomad from outside Atia (Atia II, see Map 2, sample points underlined).
It may be noted that Arabic was introduced into the western Sudanic region in the late 14th century by nomads who, at the behest of the Mameluke rulers, spread out of Upper Egypt. It is, however, noteworthy that although there are clear and specific isoglosses linking WSA Arabic and that of Upper Egypt, it is impossible to establish a dialectal link in any detail with any single dialect or dialect region of Egypt.

2.2.3 Uzbekistan and Shukriyya

Finally, two further dialects have been included in the comparison. One is Shukriyya Arabic spoken in the eastern Sudan along the Atbara River (Reichmuth 1983) and the other Uzbekistan ‘Arabic’, as described in the village of Jogari, near Bukhara (see Fischer 1961; Dereli 1997; Zimmermann 2002). There exist, unfortunately, no detailed dialect surveys of Uzbekistan Arabic. Available information points to considerable dialect variability. It is therefore important to bear in mind that what is used here is the Uzbekistan Arabic of one village, Jogari.
Arabs began settling in Uzbekistan in the early eighth century in the wake of their conquest of Transoxiana, the area north of the Oxus river in contemporary Turkmenistan and Uzbekistan. When Arab speakers in this region became cut off from the main Arabic-speaking regions is an important, though so far as I know, unanswered question. Barthold (1962) writes that in Bukhara, Arabic had ceased to be the official language as early as the beginning of the tenth century and that by this time the houses in the original Arab quarter of the city were in ruins. Uzbekistan Arabic itself has undergone such drastic influence in some grammatical domains, such as syntax (Versteegh 1984) that a case can be made for viewing it typologically as a mixed language (Owens 2001). There are thus various prima facie pieces of evidence arguing for a very early isolation from the Arabic-speaking world. As Jastrow (1998) notes, this very isolation makes Uzbekistan an interesting source for divining the form of early Arabic. The historical status of Uzbekistan Arabic plays a major role in the discussion and argumentation in §4.

For Shukriyya Arabic, there are no reliable historical accounts for when the ancestors of the Shukriyya Arabs left Upper Egypt for the Sudan or when they settled in their current location. Reichmuth (1983: 2) suggests that they have been there at least since 1800. At the same time, he concludes (1983: 29) that no direct relation with an extant Egyptian Arabic dialect is discernible, though many similarities are found with various varieties in Egypt.3

2.3 Procedure

Each linguistic variable was given as many values as it has realizations in the data. The variable 1sg perfect verb suffix, for instance, has two values, either -tu, as in qol-tu “I said” (e.g. Christian Baghdad) or -t, gil-it (Muslim Baghdad). These two values were arbitrarily given the values ‘1’ and ‘2’. The variable 1sg perfect suffix implies only the contrast -tu vs. -t. Other aspects of variation relating to this suffix, for instance whether an epenthetic vowel is inserted before the -t (as in the gilit dialect), or whether the 1sg suffix has the alternative morphophonological realization of Ø, as in Kirenawa (gul “I said”), are coded in other variables, since they are accounted for by different rules.

3. De Jong (2002: 358) considers the ‘Ababda, a group living to the north of the Shukriyya and extending into southern Egypt, to be a northern extension of the Shukriyya, influenced by Upper Egyptian Arabic.
Simple sets of statistics were produced (using SPSS) by aggregating the individual values according to variable and according to dialect groupings of different kinds. The statistical procedures I use are deliberately basic ones. The assessment of standard deviation (SD) is important, because it allows one to quickly gauge the coherency of the two dialect areas compared.

3. Statistical results and their meaning

The basic results are first presented in tables. Table 1 gives the range of standard deviations (SDs) and the means for the 49 variables. The means and SDs are calculated as follows. First, a mean and SD is calculated for each of the 49 variables individually. The individual means and SDs are then aggregated and divided by the total 49 so that an overall means/SD is obtained. Since Uzbekistan and Shukriyya have a single value, they have no standard deviation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Mean of aggregated SDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSA</td>
<td>1.68</td>
<td>.15</td>
</tr>
<tr>
<td>Mes</td>
<td>1.58</td>
<td>.39</td>
</tr>
<tr>
<td>Uzb</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>Shuk</td>
<td>1.38</td>
<td></td>
</tr>
</tbody>
</table>

The important statistic in Table 1 is the standard deviation. It is clear that WSA, with a lower SD, is a far more compact, less variable dialect than Mesopotamia. According to the measures here, it is a more coherent dialect than Mesopotamia. The standard deviation is a more telling statistic here than the mean, since it would not change, even if the values of individual variables were changed around. For instance, *k is given a value of ‘1’, *c of ‘2’. Reversing these values so that *k = ‘2’ and *c = ‘1’ would not alter the fact that WSA has no deviation at all in this variable. In the current coding system it is uniformly ‘1’, therefore SD = 0; if the values were reversed it would be categorically ‘2’, with a SD of 0.4

4. For information, 33 of the 49 variables are bivalued. Of the remaining 16, 1 has 5 values, 4 have 4, and 11 have 3. Of the 16 multivalued variables, as far as WSA goes, 10 have the same value (hence an SD of 1).
Table 2. Means and standard deviations for 49 variables, phonology and morphology

<table>
<thead>
<tr>
<th></th>
<th>Phonology mean</th>
<th>mean of SDs</th>
<th>Morphology mean</th>
<th>mean of SDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSA</td>
<td>1.72</td>
<td>.21</td>
<td>1.70</td>
<td>.10</td>
</tr>
<tr>
<td>Mes</td>
<td>1.62</td>
<td>.45</td>
<td>1.54</td>
<td>.33</td>
</tr>
<tr>
<td>Uzb</td>
<td>1.58</td>
<td></td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>Shuk</td>
<td>1.4</td>
<td></td>
<td>1.35</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Identical means, WSA and Mesopotamia compared to Uzbekistan, Shukriyya

<table>
<thead>
<tr>
<th></th>
<th>WSA</th>
<th>Mesopotamia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uzbekistan</td>
<td>19 (39%)</td>
<td>13 (27%)</td>
</tr>
<tr>
<td>Shukriyya</td>
<td>22 (45%)</td>
<td>14 (29%)</td>
</tr>
</tbody>
</table>

Table 2 repeats Table 1 except that the features are divided into phonological (N = 25) and morphological (N = 24) classes.

These statistics show that phonology and morphology contribute approximately equally to the differences between WSA and Mesopotamia. In particular, the considerably larger SD of Mesopotamian Arabic found when all features are classed together reappears in the phonological and morphological subcomponents.

Table 3 is important to the discussion in §4. Here the WSA and Mesopotamia dialect areas are compared to two specific dialects, Shukriyya in the eastern Sudan and to Uzbekistan. A count is made comparing how many times the WSA and Mesopotamian dialect areas have exactly the same score (same mean value implying same linguistic value) as Uzbekistan and Shukriyya. For instance, variable 12, absence/presence of word-internal *imala* (WSA = 1.00, Mesopotamia = 1.63, Uzbekistan = 1.00, Shukriyya = 1.00) is counted as being identical for WSA, Uzbekistan and Shukriyya.

For comparison, WSA and Mesopotomia have only 6 identical means (12% of total).

Table 3 indicates that looking beyond WSA and Mesopotamian for longer-range relatives, geographical proximity by no means guarantees structural similarity. Uzbekistan in particular is quite close to WSA, in terms of the identical means measure.

The final table breaks the Mesopotamian dialects into component parts. The justification for this is that, as seen above, the area is quite diverse. It stands to reason that this diversity would decrease in its constituent groups. The sub-
Table 4. Mesopotamian subdialects

<table>
<thead>
<tr>
<th>Dialect or subdialect area</th>
<th>mean</th>
<th>mean of SDs (where applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Baghdad qiltu (N = 2)</td>
<td>1.31</td>
<td>.12</td>
</tr>
<tr>
<td>Baghdad qiltu (N = 2)</td>
<td>1.38</td>
<td>.06</td>
</tr>
<tr>
<td>Shukriyya</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>Gilit (N = 2)</td>
<td>1.51</td>
<td>.11</td>
</tr>
<tr>
<td>Anatolian qiltu (N = 3)</td>
<td>1.61</td>
<td>.21</td>
</tr>
<tr>
<td>WSA (N = 9)</td>
<td>1.68</td>
<td>.15</td>
</tr>
</tbody>
</table>

groups used here follow traditional ones in Mesopotamian dialectology: Anatolian qiltu dialects (N = 3), non-Baghdad qiltu dialects (N = 2), Baghdad qiltu dialects (N = 2) and gilit dialects (N = 2). The present classification is based on currently accepted dialectological classification. The dialects or dialect groupings are arranged in order of increasing means. Groups with only one member have no standard deviation.

Table 4 contains surprises of different types. From a purely statistical perspective, WSA remains a strikingly coherent dialect area (has a low SD), even though it has more than three times as many members than any other group on this list and covers a much wider geographical area. It has, for instance, a lower SD than the 3 Anatolian qiltu dialects.

A more important point for present purposes is the coherency of the dialect areas as represented by the ranking of means. The qiltu dialects of Mesopotamia are split in two, as it were, not by the gilit dialects of Mesopotamia but by Shukriyya and Uzbekistan Arabic. The Anatolian qiltu dialects are closer to WSA than they are to the other Mesopotamian subdialects. Uzbekistan Arabic falls close to the middle of the entire range of mean values.

The main problem which emerges from the statistical summary in Tables 1–4 resides in the Mesopotamian dialects. They are extremely splintered, and are, under the premises set out in this study, often as or more similar to dialects geographically far removed as they are to each other. Why should this be so?

A partial answer is suggested by Nettle’s modelling of language diversification. Nettle (1999:46, 52) suggests, albeit on the basis of modeled rates of linguistic change, not actual data, that two key factors favoring language diversification are lack of population exchange between originally unitary groups, and social stratification in the population. He emphasizes that migration and social boundary favor linguistic difference. As seen in §2.2.1, Mesopotamia was pop-
ulated by two waves of Arabic-speakers, each associated with different dialects (the so-called qiltu vs. gilit dialects). Additionally, Arabs in the region have divided into socially discrete groups, the classic example being the three sectarian groups, Muslims, Jews and Christians in Baghdad (Blanc 1964) and elsewhere. The decisive role of history (different migrations) and social separation account in a broad way for the internal diversity observed in Mesopotamian Arabic. Nettle’s predictions work equally for the WSA. The dialect is comparatively uniform. To this day, Arabs have remained traditionally nomadic in this region – witness the fact that nomads included in this study from eastern Chad were interviewed in Banki and Maiduguri in Nigeria. They are also socially unstratified. They have, for instance, in their 600-year sojourn never established a ruling dynasty in the kingdoms of Kanem-Bornu or Bagirmi, within whose boundaries they live.

It is less clear what Nettle’s predictions say about how these group-internal statistics reflect on change relative to a putative, pre-diaspora ancestor. In principle, without detailed analysis of the type described in §4, the marked differences between WSA and Mesopotamia could exhibit post-diaspora development, i.e. divergence from a common ancestor, or continuation of a diversity already present in pre-diaspora times. This question may be provisionally circumvented, however, by invoking another comparison in the current data set, namely that between Uzbekistan and WSA. While too little is known about the sociolinguistics of spoken Arabic in Uzbekistan to pursue the social variable, certainly the migration variable would favor divergence, as there has been no population exchange between the areas for over 1,200 years. However, as Table 3 shows, WSA has nearly as many identical features to Uzbekistan as it does to the geographically and historically closer Shukriyya. Moreover, the features shared between WSA and Uzbekistan are found in both phonology and morphology. This suggests a relatively high degree of persistence of features, and it leads to the following hypothesis:

The statistical similarities between Uzbekistan Arabic and WSA Arabic are due to a common pre-diaspora ancestor on the Arabian Peninsula.

To be sure, this hypothesis may be criticized for not having considered the linguistic variables themselves. There are two initial answers to this.

First, the logic of the method does indeed depend on numbers for developing hypotheses and the numbers are only as reliable as the coding system and statistical test used. As far as the coding system goes, the system I used is explained at the website given above. I have tried to be as explicit as possi-
ble, but do not claim infallibility, and as noted, the statistics in fact are quite basic; I eschew the use of more advanced techniques until the basic approach is tested and the data base is expanded. Given these caveats, the hypothesis is commensurate with what little is known about the social and linguistic identity of the expanding Arabic groups. It is historically plausible in that populations of Arabs are known to have moved both into Uzbekistan and into Egypt. In the early eighth century, for instance, “Qaysites” are reported to have been settled in Upper Egypt by the Umayyad rulers (Lewis 1970:176). Qays is also a tribe reported among the soldiers of Qutayba ibn Muslim, the conqueror of Transoxiana (Agha 1999:217). Of course, the Egyptian migration is, for the WSA, a preliminary stopping point.5

It cannot be proven from the written sources at our disposal whether it was the same populations which split, one moving in one direction, one moving in the other. However, it is one strength of comparative linguistics that language can be used to elucidate earlier migrations. Unless the high degree of similarity is due to chance-independent development, it is more likely that the similar contemporary populations must at some time have shared a common ancestor.

As a second answer, I would argue that the use of statistics provides an antidote to the prevailing tendency over the last 60 years for Arabic historical linguistics to have worked with very general categories such as ‘Neo-Arabic’ or ‘Mesopotamian Arabic’. I will provide concrete examples below of problems which this approach engenders. The advantage of using statistics is that indices consist ultimately of individual cases and variables. While what comes out in the end is a single number, the number itself is dependent on individual, well-defined component parts. When the statistics quantitatively suggest a relation between two units, it is a relation defined across a range of variables.

5. It is extremely difficult to trace the exact tribal migrations out of the Arabian Peninsula, down the Nile and into the Sudan. It is, however, clear that elements who are usually reckoned to be part of Qays, at least in Egyptian genealogies, were part of these migrations. MacMichael (1967:183–184) for instance, states that the Mameluke army which made a major incursion into Dongola in northern Sudan in 1286 contained, inter alia, Banu Hilal, Banu Kanz and Rabífa. Garcin (1976:75) has identified all three of these groups as being important Qaysites who had settled in Upper Egypt. There are other tribes mentioned by name spreading in different directions. Juhayna, for instance, is eponymous for a large grouping of tribes found throughout the Sudanic region. The same tribe is mentioned spreading eastward into Iraq (Kufa) at the beginning of the Islamic conquests (Donner 1981:228). Kufa (and Basra) served as staging areas for the subsequent conquest of Iran and Uzbekistan.
At the same time, while the statistics may be suggestive of significant historical groupings, a more linguistically reliable test for measuring historical relationship is the comparative method and reconstruction of proto-forms. The nature of the comparative historical method, however, demands a painstaking reconstruction of individual forms. One does not reconstruct a dialect or an entity like ‘neo-Arabic’ but rather individual linguistic features, which are subsequently interpreted as forming larger units, such as dialects.

It is in this spirit that I consider in the next section two linguistic variables in detail which bear on the relations between the varieties discussed in this section, in particular on the relation between Uzbekistan Arabic and that of the other regions.

4. Reconstructions and interpretations

4.1 Uzbekistan Arabic

In recent years a certain amount of attention has been directed towards Uzbekistan Arabic, in terms both of its dialectological and historical status. Whereas this status has usually been discussed in terms of its relationship to Mesopotamian Arabic, with Behnstedt (2000: 145) going so far as to claim it has its origins there, I have emphasized geographically long-range associations, in particular with Nigerian Arabic (Owens 1998a: 72).

The statistics in Tables 1–4 would support both associations. However, Behnstedt’s peremptory dismissal of the Nigerian connection in favor of a Mesopotamian one cannot be upheld in the face of Tables 3 and 4: Uzbekistan Arabic has more variables identical to values of WSA than to Mesopotamia, and its mean values, though globally considerably closer to Mesopotamia than to WSA (Table 1), does not align with a single Mesopotamian subgroup (Table 4).

An initial conclusion is that the present statistical comparison does not contradict my suggestion of a significant relationship between WSA and Uzbekistan. However, the statistical summary is neutral as to historical interpretation. Similarities and differences may be due to common inheritance or to shared or divergent innovations. Given that the current set does comprise fundamental phonological and morphological features, it does give a broad basis of comparison. The WSA is testimony to this. Its relative uniformity can plausibly be related in part at least to its historical roots and social character (see
§3 above). By the same token, the fact that the WSA exhibits a good number of similarities to Uzbekistan encourages a closer look at the two varieties.

Turning to this question, Behnstedt says that Uzbekistan Arabic has its origins in Mesopotamian Arabic. To this point I have not treated historical linguistic questions in detail, so in a sense, what I have presented and what Behnstedt asserts have no direct connection with each other. However, Behnstedt himself does not show how Uzbekistan Arabic developed from Mesopotamian, so his assumption of direct genetic affiliation is unsupported.

In making his claim, Behnstedt refers to Jastrow’s work (without a specific reference, however). In fact, Jastrow (1998) does treat the relation between Uzbekistan Arabic and Mesopotamian Arabic in historical linguistic terms. Jastrow begins by noting that in looking for relationships between Mesopotamian Arabic and dialects outside of the area, one should not consider Mesopotamian Arabic globally. Rather, he divides it into the familiar qiltu and gilit dialects, and proceeds, initially at least, in the same way data has been organized in this paper. He draws historical inferences on the basis of contemporary distributions of variables, though with considerably fewer than in the current data set. In the following Table 5, Jastrow’s data (1998:177) is reproduced in the first three columns, representing Uzbekistan Arabic, qiltu dialects (Q) and gilit dialects (G). I have added my own data from Nigerian Arabic. The variables used in the comparison are copied from Jastrow. The ones identical to those used in my statistics are identified by the variable number in my data (see §6). I have added one variable to Jastrow’s list, namely the realization of qaaf, as this is, traditionally, one of the constitutive variables distinguishing the qiltu and gilit dialects.

Below the table I have counted how many features of similarity there are between Uzbekistan Arabic and the other varieties. This simply summarizes Jastrow’s own presentation and is not intended to add another statistically-representative data set to the discussion.

Even more strikingly than in my own data, this abbreviated list confirms that a prima facie case can be made for linking Nigerian Arabic with Uzbekistan. It furthermore confirms the heterogeneous nature of the Mesopotamian area; the gilit dialects in this reckoning have only two features in common with Uzbekistan Arabic.

At this point it is time to move beyond tabular listings. Methodological clarification is needed, however. For Jastrow, Table 5 is not so much a taxonomic listing as a statement about historical relations. He says that his list is based on “archaic characteristics” (“altertümliche Merkmale”). The idea is that
Table 5. Uzbekistan Arabic compared to other dialects: 7 variables

<table>
<thead>
<tr>
<th></th>
<th>Uz</th>
<th>Q</th>
<th>G</th>
<th>Nig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfect endings</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>-iin, -iun (6.2.52)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>M vs. F plural</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>(6.2.30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linker -in</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>2s.sg -ki (6.2.62)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>internal passive</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>1sg perfect -tu</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>(6.2.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>qaaf (6.1.1)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Shared traits: Uz – Q: 4/7, Uz – G: 2/7, Uz – Nig: 5/7

the variables listed in the comparison represent features going back to Old Arabic or Proto-Arabic (see below), or at least to a stage of Arabic before certain innovations occurred in the Mesopotamian dialect area.

As it stands, however, Table 5 is no more a statement of historical relations than are the 49 variables in my data bank. Lacking an explicit demonstration of which features on the list are old or proto-forms, one runs the risk of claiming or assuming one feature to be older than another in a comparative linguistic sense, while in fact there is no linguistic basis for the assumption. A case in point pertains to a further element in Jastrow’s presentation.

4.2 New is not new until proven new

Jastrow seeks to demonstrate that the Mesopotamian qiltu dialects have collectively undergone certain innovations. This explains why Mesopotamian qiltu dialects differ in certain respects from Uzbekistan Arabic: whereas Uzbekistan Arabic has retained archaic features, the qiltu dialects in their post-diaspora phase innovated.

A case in point is the well-known phenomenon of imala (1.12). Very briefly (see below), this is a type of vowel harmony, where (in the word-internal type), a long aa changes to ie or ee in the presence of an i in a neighboring syllable.

Imala is one of the features Jastrow cites as representing a post-diaspora innovation characteristic of the Mesopotamian qiltu dialects. He uses this observation to argue that Uzbekistan Arabic represents a “type” that “...did not par-

6. See Owens 1998b; Ferrando 2000 for discussion of this morpheme. The internal passive (e.g. iktib “it was written” vs. kitab “he wrote”) in the Mesopotamian gilit dialects in fact is quite variable. Many, like MB and some southern Iraqi dialects, do not have it.
ticipate in either of the waves of innovation of the two [dialects, i.e. qiltu/gilit – JO]; it displays the innovations of neither the qiltu nor the gilit dialects ...” [my translation]. Jastrow further notes that Uzbekistan Arabic in these points, largely agrees with Classical Arabic.

Imala, however, is not a post-diaspora innovation and therefore the qiltu dialects cannot be said to have innovated the feature. This can be seen on two counts. First, imala is described in great detail in the oldest detailed source which exists for Classical Arabic, namely Sibawaih’s Kitaab (II: 279–294). Sibawaih presents it as an alternative to non-imala forms, where apparently the non-imala forms are less characteristic of Hijaz Arabic, imala more common in the eastern Tamim region (II: 280, 282). Nowhere does Sibawaih differentiate the imala or non-imala variants in terms of prestige, correctness, or other evaluatory criterion.7 Levin (1998) has summarized the linguistics of imala in Sibawaih, and I will not elaborate on his description of its governing conditions here, beyond noting one point. The conditioning factor in imala, whether in the modern dialects or in the classical language, is essentially the same. Long aa changes to a sound like ie or ee in the presence of an /i/ in a preceding or following syllable. For instance, in eastern Libyan Arabic the /aa/ of perfect verb saamah “he forgave” therefore is “changed” to /iel/ in the imperative or imperfect, due to the presence of the following /i/, siemih “forgive!”. Sibawaih describes the same change (‘ielim, “learned, scholar”, II: 279).

To underscore the basic identity of the process one may consider what is sometimes in Arabic dialectology regarded as irregular imala because it appears not to be conditioned, as in nies “people” > naas. This is not, contrary to Jastrow (1978:66), fundamentally different from classical imala. Precisely the

7. Levin’s statement that imala was not a general phenomenon in “old Arabic”, but found only in some of the ancient (i.e. eighth century) dialects begs the question of what “old Arabic” was (2002:444). In this passage he appears to hold the view that Sibawaih neatly divided Arabic into dialects and Standard (the Arabiyya), and that imala falls into the former category. Such an explicit dichotomy was not drawn by Sibawaih as regards to imala, nor for other variational phenomena he treated. Elsewhere Levin (1998) correctly notes that imala is an accepted variant, both among grammarians and in Quranic recitation. In this same article he already points out that classical imala and that found in the modern Mesopotamian dialects are conditioned by very similar factors and he goes on to suggest a historical connection between them.

8. This is an Eastern Libyan form. Note that in the very last chapter of his Qiraa’aat (p. 703) Ibn Mujahid mentions that Kisaa’t used imala in the genitive of the word naas = n-nies-i.
same word, *nies*, is mentioned in Sibawaih as undergoing *imala* (II: 285), where he describes its *imala* as exceptional (*faadīd*).

*Imala* in the Old Arabic tradition is moreover reported to be at least a generation older than Sibawaih. In the *qiraa‘aat* literature we find in particular that Kisaa’ī (Sibawaih’s contemporary) and Abu ʿAmr Ibn ʿAla‘ā (d. 770/154, one of Sibawaih’s teachers) were two readers who used *imala* regularly, both word-finally and word-medially (Ibn Mujahid 1972: 146–152).

Now, there exist no written sources for Arabic older than Sibawaih and the *qiraa‘aat* tradition for the phonetic details required for an understanding of *imala*. *Imala* is thus as much a part of the Classical language as are non-*imala* variants. Furthermore, the reading tradition as summarized by Ibn Mujahid (d. 936/324) contains the *Quran* recitation practices of seven readers who flourished between 738 (Ibn ʿAmir d. 738/118) and 804/189 (Kisaa‘ī). It may be assumed that *imala* was not innovative with them, but rather that they used a phonological trait whose origin is older, as yet undated in a comparative linguistic sense. Arabs began settling in Uzbekistan (Transoxiana) after 710. In terms of chronological time, therefore, *imala* is attested at a date contemporary with the Arabic settlement of Uzbekistan. For whatever reasons Uzbekistan Arabic does not have *imala*, one can be excluded, namely that *imala* did not exist at the time of the original Arab settlement of the country. It is not the case that it was only after Arabs were comfortably ensconced in Transoxiana that elsewhere in the Arabic-speaking world *imala* emerged.

This is the first reason *imala* cannot be regarded as innovative in the *qiltu* dialects. The second argument relates to the contemporary distribution of *imala* in Arabic dialects. Though often ignored by Arabicists working on the phenomenon, a productive word-internal *imala* is also attested in Eastern Libyan Arabic and in Malta (Owens 1993c). What is important is that the distribution of the same phenomenon in such widely separated areas as Eastern Libya and northern Mesopotamia points to a single common point of origin. *Imala* is historically contingent. The simplest explanation is that *imala* of the present-day dialects originated in pre-diaspora Arabia, and spread from there in different directions. This indicates a pre-diaspora origin, i.e. one which can be reconstructed no later than the seventh century.

9. Ibn Kathiyr, Naafī, Hamza, Ibn ʿAmir and ʿAasim on the other hand are mentioned as those who do not generally use *imala*, though all may use it occasionally, e.g., p. 151. The Kufan reader Hamza is said to use *imala* in *ʔaimata* (*ʔamieta*) “who granted death”, (53:44), “stronger even than that of Naafī and Abu ʿAmr”.

---

Pre-diaspora Arabic 293
The dating of *imala* impinges on the present discussion in the following way. Jastrow has argued that Uzbekistan Arabic is in some sense “older” than the *qiltu* and *gilit* dialects of Iraq in that the Iraqi dialects have innovations not found in Uzbekistan. Looking closely at how this one ‘innovation’ is characterized, however, shows that *imala* is a variable which in fact cannot be used to establish the relative age of Uzbekistan vs. Mesopotamian *qiltu* Arabic, since *imala* can be reconstructed into pre-diaspora Arabic and therefore predates both Mesopotamian and Uzbekistan Arabic. Similar problems arise with various other features which Jastrow wants to establish as touchstones for innovation (1998: 178), though this is not the place to go into further detail.\(^{10}\) The crucial point is a methodological one. Whereas Jastrow interprets the list in Table 5 in terms of historical change, without an unequivocal standard by which to measure the change, no historical linguistic interpretation is possible.

At this stage in the study of Arabic, questions of what is innovative and what is old or retentive have to be answered on a case by case basis. While I would assume consensus on some features, for instance, that a proto-language had a distinct *f.pl* morphology, on others there is none, or, as I have shown in the case of *imala* there are many features which Arabicists might have assumed to uncontroversially be divisible into an old/innovative dichotomy, which on closer inspection are not amenable to such a division at all.

The fact that in the present sample the *gilit* dialects, Uzbekistan and WSA have no *imala*, would constitute one small indication of very old affiliation between these varieties, despite the great geographical distance separating them.

4.3 What is not attested in writing is not necessarily non-existent

As I noted, the statistics in §3 are useful as general direction markers. Besides the need to establish the historical depth of each feature, as illustrated in §4.2, shared features which are relatively rare are a valuable diagnostic for establishing historical relationship. Thus, the fact that the form *gul* “I said” cited in §2.3 is shared among all WSA dialects is significant not only for its limitation to the WSA, but also in its status as a morphologically unusual linguistic feature. Such unusual events are unlikely to be produced more than once.

Interesting in the present context is a feature shared between Uzbekistan Arabic and some WSA dialects (the Bagirmi dialects) whose chance of indepen-

\(^{10}\) Consider for instance rules relating to syllable structure and affrication of *k* and *g*.
dent origin is quite small (not included in the sample for lack of exemplification in some works consulted).

Roughly speaking, in Arabic dialects there are three forms of an active participle with pronominal object suffixes. The three different forms are contrastive in the F.SG. In all varieties the non-suffixal feminine takes the regular F.SG adjectival suffix -a or -e, the vowel quality depending on harmony rules, if any, prevailing in the local dialect.

Turning to the object suffix forms, in one widespread alternative (e.g. Cairene, non-Bagirmi WSA dialects) the object suffix is suffixed to the F.p articipial form marked by -ee, or -aa, which otherwise is the F.SG adjectival marker (see (4a)). In a second, when a pronominal object suffix is added to a feminine form the feminine participle takes the morphophonological alternative -it, otherwise used in a genitive construction (see (4b)). This is found for instance in Eastern Libyan Arabic and in many Arabian Peninsula dialects. The third alternative is to add an intrusive suffix -in- or -inn- to either the M.SG or F.SG form whenever a pronominal object suffix is added, as in (4c)).

The third alternative (4c) is by far the least common. It is found only in relatively small, isolated areas, so far reported only in Oman and western Hadramaut, Bahrain (among Shia Baharna), Uzbekistan, Khorasan (Seeger 2002: 635) and Bagirmi Arabic in the WSA. There are differences between its form and use in these areas. I will limit myself to an exemplification of three manifestations here. In Uzbekistan Arabic the first pronominal object suffix added to an active participle marks the subject of the sentence, as in (5a).

In all other areas only one pronominal object is allowed, and it marks the object. In Oman the intrusive -in- is suffixed to the gender/number markers of the participle:

a. zorb-in-naa-kum
hit-in-we-you.m.pl
"we have hit you".

b. ðaarb-it-n-ij
hit-f-in-you.f.sg
"she has hit you.f" (Reinhardt 1972: 139)
In Bagirmi the intrusive *-in* is suffixed directly to the participle stem, thereby neutralizing gender and number contrasts, *(hi) d˚aarb-in-he* “(she) has hit her”, identical to masculine, *(hu d˚aarb-in-he)* “(he) has hit her”.

What is important for present purposes is the observation that it is extremely unlikely that so unusual and specific a feature as the intrusive *-in* could have originated independently in four or five geographically separated areas. It is plausible to think in terms of common place of origin, with the present-day geographical distributions being accounted for by migration out of this place of origin. Following Retsö (1988: 88), a common origin, however, could only be somewhere on the Arabian Peninsula, so far as the present-day evidence allows us to deduce, on the eastern and northeastern fringe, and it would have had to have been pre-diaspora, i.e. pre-640 C.E. Pre-diaspora Arabic, however, is contemporary with what is customarily known as Classical Arabic. To my knowledge, the intrusive *-in* is mentioned nowhere in Classical Arabic, either by a grammarian or in a *Quranic* reading tradition. It does not exist as a written, attested form. Nor can it be derived via grammatical rule from a Classical Arabic form or forms, as Retsö (1988) shows. Yet simple principles of reconstruction, as briefly described here, require its presence as a proto-form in the seventh or eighth century. It is a contemporary of all varieties mentioned in early written sources. Clearly then, Classical Arabic as described in old sources does not exhaustively describe the forms of Arabic which were spoken during its era. The comparative method based inter alia on a consideration of modern dialects forces us to reconstruct further forms in that era.

4.4 Innovations and retentions: Western Sudanic Arabic

The detailed case studies from the previous two sections serve as models for a broader interpretation of Arabic linguistic history. Both of them point to forms prevalent in peninsular Arabic in pre-diaspora times, and both point to a dialect differentiation present at that early era. *Imala* and non-*imala* forms, for instance, were clearly coexistent. The broader implication is that diversity found today among the Arabic dialects can mirror diversity already present in pre-diaspora times.

It is clear, of course, that there have been innovations, sometimes striking ones, since the Arab diaspora of the seventh and eighth centuries. Pinpointing them, however, is more problematic than sometimes assumed in Arabic linguistics (see §4.3). It is beyond the scope of the present paper to look at each of the 49 features treated here on a case by case basis, though such de-
tailed treatment is necessary. I believe that only the following elements are self-evidently post-diaspora innovations in Western Sudanic Arabic. I will not treat Mesopotamian Arabic here, for as seen it has a more differentiated history than does WSA.

(6) Possible and likely innovations in WSA
(numbers reference the features in §6):

Phonology (features 6.1.3/ 4/ 7/ 8/ 9/ 10)
- h > h, ñ > ñ, δ > ð/ð, ð > t/s, ð > d˚, t˚ (?) > ñ.

Morphology (features 6.2.44/ 53/ 54/ 72)
- Loss of -t 1/2m perfect verb suffix in certain contexts, n- "I" -u "we"
in imperfect verb (Owens 2003 for these two forms), 3r.sg form of weak imperfect verb.
- All other features included in this data bank, I would argue, are candidates for pre-diaspora Arabic. Note that I include in this latter set forms which at some point would disappear into deeper Proto-Arabic forms. Variable 6.1.6, for instance, is the reflex of Standard Arabic γ. In WSA this has various reflexes, though the most common is q attested in 6 of 9 sample points. In the present sample the reflex q is also attested in Shukriyya (especially in final position) and in southern Mesopotamia. Outside these areas it is found among other places in the former North Yemen (about 5 sample points in Behnstedt 1985:44), in Syria in a long, continuous area on the Euphrates River, stretching to the Turkish border, and westward to Aleppo (Behnstedt 1997:16) and in Hassaniyya of Mauritania. Assuming this reflex did not innovate spontaneously over and over again, the broad distribution of the variant speaks for a single, pre-diaspora origin. At an older pre-diaspora level the innovation γ > q occurred among some groups of speakers. Populations with both variants then moved outside of the peninsula.
- The important point, however, is that the broad distribution of q < γ outside of Arabia speaks for a reconstruction of the form in the pre-diaspora population as well. All features not included in (6) above, 39 in all, are candidates for a pre-diaspora variety, though of course ultimate justification depends on individual argumentation. I would note that none of the innovations in (6) is shared between WSA and Uzbekistan Arabic.
5. The interpretation of Arabic linguistic history

The discussion in the previous three sections may be seen as complementing the use of the statistics in the following way. The statistics provide, as it were, a blunt instrument for suggesting where significant relationships lie. As raw data, they do not automatically reveal why dialect areas are similar or different, whether because of chance convergence, change due to contact and borrowing, innovation due to simplification or analogical formation, or to common inheritance. They can, however, help identify where it might be interesting to look for further evidence which helps disentangle what elements are in play. Thus, by one statistical measure of similarity and difference (Table 3), Uzbekistan Arabic turns out to be as similar to WSA as it is to Mesopotamian Arabic. These similarities, moreover, are most likely not due to independent innovation (see 4.3). Looking further to features shared by these two varieties it turns out that both have a highly contingent participial construction, whose origin goes back to pre-diaspora Arabic. The shared nature of this feature can only be explained in terms of an original pre-diaspora population with this feature in their language moving out of the Arabian Peninsula, then splitting, one eventually settling in the eastern extremity of the Arabic-speaking world, the other at an extreme western end. In this case, statistics and reconstruction of the participial forms complement each other vividly.

The statistics, furthermore, provide the basis for other interesting, if not linguistically powerful, hypotheses. I have, for instance, included Shukriyya Arabic only as a control to provide a counterfoil for the emphasis on Uzbekistan Arabic. The statistics in Tables 3 and 4, however, indicate quite cogently that Shukriyya has no noteworthy affinity with WSA as compared to Mesopotamian Arabic. I do not think this is statistical trivia. Rather, it indicates that the form of present-day Arabic dialects (as, of course, measured in the current 49 variables) in some cases can be as or even more dependent on inheritance than it is on geographical proximity. In dialectological terms, there is, in fact, no strong evidence here for a ‘Sudanic Arabic’ (as I term it) contrasting as a whole with Mesopotamian.11 The historical linguist is challenged here on two fronts: where, historically, did the differences come from and to what extent does coterritorial contact hinder or abet dialect differentiation?

11. Even less so Behnstedt’s (2000:145) facile categorization of Sudanic Arabic as Hijazi (Saudi Arabia).
Lurking behind the discussion of the individual variables used in the statistics and reconstruction of specific linguistic forms is the more fundamental issue of how, generally, Arabic linguistic history is to be conceptualized.

Essentially there are two interpretative approaches. One, developed in Germany in the nineteenth century and hardly altered since, is to regard Classical Arabic, conventionally termed ‘Old Arabic’ (Altarabisch) as the ur-ancestor of the contemporary dialects, also known as ‘Neo-Arabic’ (Neuarabisch, e.g. Brockelmann 1908; Fischer 1995; Blau 1988). Jastrow’s model follows this tradition. He assumes a division between old and new or innovative structures, and on this basis works out a historical linguistics. Thus, imala is assumed to be innovative or newer than non-imala varieties, and so Uzbekistan Arabic, which lacks imala, can be considered an older variety than are the qiltu dialects of Iraq, which have the innovative imala. This model has the advantage that the work of linguistic reconstruction, normally the sine qua non of historical linguistics, is obviated through the presence of an already defined Classical Arabic.

The problem with this approach is not in its basic logic. In some respects Jastrow’s interpretations are certainly correct. The change of $r \rightarrow y$, characteristic of a few Iraqi qiltu dialects (see Appendix 6.1), is indubitably innovative for those dialects which have it. The problem lies in the assumption of how the ancestral variety is determined.

A second approach is to assume no predefined ancestral version, and to develop one according to customary principles of comparative linguistics. In this perspective, Classical Arabic is included in the consideration of any reconstruction, but cannot automatically be assumed to have a privileged position relative to other varieties, even contemporary ones. An analogy can be drawn with Indo-European. Whereas Sanskrit and Hittite are the oldest attested examples of Indo-European, neither is assumed to be identical with Proto-Indo-European, which is reconstructed in part on the basis of varieties attested at a more recent period than Hittite or Sanskrit.

Ferguson’s well-known proposal (1959) that the Arabic dialects arose in a military-camp koine I regard as a specific variant of the Old/New dichotomy, an attempt to give a social explanation to an assumed historical dichotomy. As subsequent dialectological research has shown, the individual reflexes which Ferguson proposed were evidence of a uniform koine, opposed to Classical Arabic, have proved far too diverse to have originated in a limited number of military camps. Kaye (1976) is an early criticism of Ferguson’s koine hypothesis, a criticism which should be updated in light of intervening research on Arabic dialects.
The second position is little developed among Arabicists, and so what I have presented here represents only basic spadework. It leads to a reconstruction of pre-diaspora Arabic which is considerably more complex than traditionally assumed. The complexity in part, as seen in the discussion of Sibawaih’s *imala*, is already attested in some detail in the oldest texts, and in part it follows from a simple reconstruction of forms based on the distribution of post-diaspora elements, such as the intrusive *-in* treated in §4.3. It can also lead to the postulation of forms which may even contradict aspects of Classical Arabic as described by the Arabic grammarians (Owens 1998b). The cost of a more linguistically adequate characterization of Proto-Arabic will involve, in the short run at least, the juxtaposition of sometimes incompatible forms.

The problem treated here is not new. Indeed, it was summarized over 1,000 years ago by Ibn al-Nadim (d. 954) in his *Fihrist*: “. . . and each [pre-Islamic – JO] tribe had its linguistic variety which distinguished it and by which it was recognized, all of which are a part of the original (*asšl*) variety”.

The problem of diversity is thus not only not new, but also all-pervasive. Summarizing research on a collection of old Quranic fragments found in San’a, Yemen, Puin (1996: 109) writes, “. . . we discovered that ‘our’ manuscripts contain many more Qiraa’aat than are recorded by the old authorities”. Whether and in what sense these fragments are older, as Puin suggests, than the traditional ‘seven’ version referred to at the beginning of this article is a question requiring greater historical and philological scrutiny. Clearly, however, even in the domain of early written Arabic the canonical Quranic reception, itself already variegated, tends to understate the degree of diversity in the language.

In a recent compendious survey of the term “*tarab*” and its congeners in classical sources Retsö suggests that the Classical language itself, as embodied in pre-Islamic poetry, had died out before Islam, and that it existed not in

---

13. The two general and opposing positions which I define here are, in the contemporary state of Arabic linguistics, relatively poorly profiled. In fact, only the first has much currency. This is unfortunate, as I believe it may be associated with a highly scholarly, but at the same time orthodox and restrictive interpretation of Arabic linguistic history. Among its best known representatives are Brockelmann, Nöldeke and Fück. What today is too little appreciated is that contemporaries of Brockelman and Nöldeke such as Völlers, de Landberg and later Kahle (see Fischer 1982 for discussion) argued for a broader reading of what the Arabiyya was. Without agreeing with all interpretations of this latter group, I would see my position as reviving their perspectives.
one form but rather as "several 'languages of the Arabs' in pre-Islamic Arabia" (2003:595).

Fleisch draws a specific connection between social organization and dialect diversity when he writes (1974:13) that "On se trouve donc devant une fragmentation de la population, poussée à un très haut degré; la conséquence naturelle est la fragmentation du langage: le dialecte". Fleisch, however, works within the 'Old Arabic – Neo-Arabic' paradigm, whereby old diversity is lost in the course of post-diaspora simplification.

Against Fleisch, the thrust of this paper suggests that old diversity is often directly reflected in contemporary diversity, and that contemporary diversity can be used both to reconstruct old diversity, and to explain contemporary diversity through the ultimately simplifying procedure of the comparative method. This can take place even in the absence of confirmation in old, classical sources. From this vantage point, modern scholarship has a further tool to discern what the ḥaǧl of Arabic is.

6. Appendix: Summary of variables

In the text these variables are referred to under 6.1 for phonology or 6.2 for morphology plus the number in brackets, which is the number by which the variable is identified in the data bank. /k/, for instance, is 6.1.2. The numbering sometimes has gaps in the sequence, these being left for related features.

6.1 Phonology

(1) ‘qaf’ (reflex of SA /q/); 1. q, baqar “cattle”: Mardin 2. g, bagar: Kirena.
(2) /k/ 1. k: kalib: CB, 2. c [t]: caib: MB.

In those dialects with a k ~ c alternation, c basically occurs in the environment of front vowels, or high front vowels but not back ones. The precise conditions vary from area to area, however.

(3, 4) /h/ /ʃ/: 1. laham, “meat”, baa’at, “she sold”; 2. laham, baa’at: Kirena.
(5) *ha/’a; 1. laham, ga’ad “he stayed”, 2. lehem “meat”, *ge’ed: Bagirmi.

In the Bagirmi area proto *ha/’a raise to *he/’e. Subsequent to this raising of /a/ to /e/, (3, 4) above occur.

(8) /θ/; 1. θ, ḥaθa “three”: Hiit, 2. t, ṭalataa: Kirena, 3. s, saḷ: Aajiri, 4. f, faafe: Siirt.
This correspondence applies to basic vocabulary.

(9) /ð \ddot{o}/; 1. ð\ddot{o}, ð\ddot{a}all “remain”: Hiit, 2. d\ddot{o}, d\ddot{a}har “back”: Kirena wa, 3. ð\ddot{a}: Daragözü, 4. ð\ddot{a}h: Siirt.

(10) t\ddot{a}/; 1. t\ddot{a}, t\ddot{a}r “it flew”: JB, 2. d\ddot{a}, d\ddot{a}r “back”: Kirenawa, 3. t\ddot{a}, t\ddot{a}r: Abbeche.

(11) Realization of “jiim” 1. j, jimal “camel”: (MB), 2. y, yimal: (southern Mesopotamia).

(12) Word-internal imala 1. No imala, kaatib “has written”: Kirenawa, 2. Imala, keetib: CB.

The imala in the Mesopotamian area has two degrees, with some dialects (e.g. JB) having a further raising to ii in some forms (e.g. nominal but not participial, kliib “dogs”).

(13) r\ddot{a}; 1. r\ddot{a}, baraid “cold”: Kirenawa, 2. r\ddot{a}, bee\ddot{i}: CB.


Jastrow (1973:27) considers the stem /a/ in the Daragözü form t\ddot{a}fa “you, f. open” to be an epenthetic vowel. This creates an interpretive problem. In terms of underlying form this would be classified in category “3”, but in surface realization as (1). I have classified it as “3”.


(17) CaCii; 1. CaCii, kabir: Kirenawa, 2. CCii, kbiir: Hiit, 3. CiCii, ci\ddot{i}: MB.

Sasse (1971:238) notes that both (1) and (2) occur in Mardin, nab\ddot{i} “clean”, gbiir “big”, on a lexically governed (i.e. irregular) basis.

(18) CaCa(C); 1. CaCa(C), katab “he wrote”, bogar “cattle”: Abbeche, 2. GiCa(C), citab, bugar: MB.

(19) CaCa-C (verb); 1. CaCa-C, katab-at “she wrote”: Kirenawa, 2. CiCC-V, kitb-at: MB, 3. CaC-C: ktab-it: JB.

(20) CaCaC-C(V) (verb); 1. CaCaC-C, katab-tu “I wrote”: Mardin, 2. CaC-C, ktab-it: Hiit, 3. GiCaC-C, ktab-t: MB (or ktab-t).


(22) CVCC# (noun); 1. CVCC, xxt “sister”: JB, 2. CVVC, uxut: MB.


(24) CguCa gahwa (guttural = x, y, q < y, ı, h, h, ğı); 1. CguCa, gahwa: Mardin, 2. CguCa gahwa: Kirenawa.

(26) CvCVC(C) stress; 1. CvCVC, bugar “cattle”: Kirenawa, 2. CvCVC, bu\ddot{a}gar: Aajiri.

(27) Emphasis; 1. Emphatic consonants, t\ddot{a}y “it flew”: CB, 2. No emphatic consonants, taar: Abbeche.

6.2 Morphology

(30) feminine plural; 1. yes, buktub-an “they, f. write”: Kirenawa, 2. no, ykitb-u (common plural): MB.

(31) two verb conjugations, low vs. high vowel (perfect verb); 1. yes, katab-tu /kb\ddot{a}y -tu “I wrote/I grew”: CB, 2. no; ktab-it “I wrote”, lbas-it “I wore”: MB.
The two conjugations are characterized by an opposition between a high stem vowel and a low stem vowel.

(40) First person singular perfect suffix: 1. -tu, katab-tu “I wrote”: CB, 2. -t, katab-it: MB.

I have not included the variant -eet which occurs in some southern Mesopotamian dialects (e.g., Shatt al-Arab), kitb-eet “I wrote”.

(41) 3 sg. perfect suffix: 1. -at, katab-at “she wrote”: Kirenawa, 2. -it, katab-it: Mardin.


The variant -taw found in southern Mesopotamia is included under value “4” here. In Daragözü -to is minimally contrastive with the 1sg perfect suffix -tu.


In some qiltu dialects, before a pronominal suffix the variant -aw may appear, katab-o ∼ katab-aw-ha.

(44) 1sg perfect suffix; -t or (t); 1. always -t (or -tu), katab-tu: Mardin, 2. -t morphologically conditioned, katub “I wrote”: Kirenawa.

The conditioning factor for (2) is complicated (see Owens 1993a:104).

(45) verb C-aa-object suffix; 1. no, katab-ha “he wrote it”: Kirenawa, 2. yes, katab-aa-ha, jaaf-aa-ha “he saw her”: Abbeche.

(51) preformative vowel a, i or a/i; 1. a: yaktub “he writes”, 2. i, yiktub: Kirenawa.

There are no dialects in this sample with only /a/ (southern Borno Arabic does have such). If there are two values, a, i, the quality is usually determined by harmony with the stem vowel. Other distributions are possible, however. In Umm Hajar in Chad, for instance, i occurs in open syllables while a occurs in closed, t-i-bii “she sells”, t-a-ktub “she writes”. Dialects with both i and a have a value of “1.5”.

(52) imperfect suffixes (without object suffixes); 1. -uun, -iin, yiktub-uun “they write”: MB, 2. -u, -i, yiktub-u: Kirenawa.

Siirt has a special stressed form, without -n, yonk_hover “they were broken”. I have coded this as 1.5, since it lacks -n (coding = “2”), but in its special stress is differentiated from all other dialects which do not have -n. Alternatively, one could give it a separate coding of “3”. Before suffixes, in some dialects (many Anatolian qiltu) the n is deleted, Mardin yiktub-uu-ha “they write it”: I do not usually catalogue variants of suffixes with a further attached suffix.

(53) 1 singular imperfect; 1. a-, a-ktub “I write”: Kirenawa, 2. n-, n-uktub: Abbeche.

The 1sg prefix is a-. In the indicative it will often be prefixed with b-, ba-ktub.

(54) 1 plural imperfect; 1. n-, n-uktub “we write”: Kirenawa, 2. n- ...-u, n-uktub-u: Abbeche.

(61) 1 singular object suffix on noun 1. -i (unstressed), 'beet-i "my house": CB, 2. -ii, 'beet-ii: Abbeche.

Nigeria is distinguished from Chadian Arabic in that in Nigerian Arabic the verbal object suffix -ni is not stressed, whereas it is in Chadian. This difference is not catalogued.

(62) 2 r.sg. object suffix; 1. -ik, 'beet-ik "your house": CB (after C-), 2. -ki, 'beet-ki: Kirenawa.

A number of dialects (e.g. CB) have both variants on a conditioned basis, e.g. -ki after a long vowel, otherwise -ik.


I do not classify according to the vowel of this suffix, which can be distinctively front or back. In Hiit it is in fact [-him]; similarly, for the preceding 2m.pl. suffix.

(64) 3pl. object suffix; 1. -hum, 'beet-hum "their house": Hiit, 2. -hun, 'beet-hun: Mardin

(65) 3 m.sg. object suffix; 1. -u, 'beet-u "his house": Hiit, 1.2. -u, VV-nu, katab-uu-nu "they wrote it", CB, 1.3. -u, VV-hu, jiib-uu-hu "they bring him": Mardin, 2. -a, 'beet-a: Kirenawa.

Bagirmi Arabic has -e after a front vowel ('beet-e), otherwise -a. This variable is an exception to my practice of not treating post-suffix allomorphy. However, the allomorph u ~ nu ~ hu is defined by any vowel, not only a suffix vowel, e.g. safaa-nu "his supper". Variants with the -nu ~ hu post-vocalic allomorph are given a % classification which is closer to "1" than to "2", since the component vowel -u is closer to -u than to -a. After a vowel, dialects classified here as "1" or "2" both have the variant ð (stress shift and length), e.g. sa'aaa "his supper" (both Hiit and Kirenawa, allowing for pharyngeal change).

(66) H-deletion from object pronouns; 1. no h-deletion: 'beet-hum "their house": Khaweeta, 2. h-deletion: 'beet-um: Kirenawa.

There is a large amount of variability in the deletion of the /h/, e.g. in Daragözü in Turkey the /h/ has apparently completely disappeared, so that even after a vowel only the h-less form occurs, e.g. kol-tuu-a "I ate it.s". This variant has not been given a different coding.

(67) Stress shift before V-initial object pronouns; 1. no shift, masak-u "he grabbed him": Kirenawa, 2. shift: ma'sak-u: Mardin.

(68) Form of 2m.sg object pronoun; 1. -ak, 'beet-ak "your house" (MB), 2. -ik, 'beet-ik (Siirt).

(71) Doubled verbs 3m.sg; 1. CaCCa, tamma "he finished": Kirenawa, 2. CaCC, tamm: CB.
Pre-diaspora Arabic

(72) weak final verb, vowel suffix; 1. -Vt, -o/u, nis-at “she forgot: Kirenawa, 2. (i)y-Vt, nisy-at: Mardin.
(73) Initial vowel of imperfect of aCaC verbs; 1. prefix + aa, yuaxud “he takes”: Kirenawa,
2. prefix + oo, yooxud or yooxuz; Uzbekistan.

Abbreviations and Dialects used in survey (number represents the code)
JB = Jewish Baghdad Arabic, CB = Christian Baghdad, MB = Muslim Baghdad, WSA =
western Sudanic Arabic; for Mesopotamia, /i/ = undifferentiated high, short vowel.
Mesopotamia (dialect subgroup as in Table 4 given in brackets): 10. Christian Baghdad
(Baghdad qiltu), 11. Jewish Baghdad (Baghdad qiltu) 12. Muslim Baghdad (qilti), 13. Mardin
qiltu), 18. Hiit (non-Baghdad qiltu), 19. southern Mesopotamia (qilti)
30 Uzbekistan, 31 Shukriyya

References

46.211–229.
Barthold, Vasily. 1962 [11928]. Turkestan Down Through the Mongol Invasion. London:
Luzac.
Blanc, Haim. 1964. Communal Dialects in Baghdad. Cambridge Mass.: Center for Middle
Eastern Studies.
Brockenmann, Carl. 1982 [11908]. Grundriss der vergleichenden Grammatik der semitischen
Sprachen. Hildesheim: Olms.
Wiesbaden: Harrassowitz.
University.
al-Farra’, Abu Zakariyya. Ma‘aamiy al-Qur‘aan. ed. by Mohammad Al-Najjar & Ahmad


Résumé

Les dialectes arabes s’étendent sur une vaste région, à savoir, de l’Iran à la mer du Tchad et du Yémen au Maroc. Il en résulte une considérable diversité linguistique. L’explication de cette hétérogénéité est toutefois problématique. À cette fin, quatre dialectes arabes nettement différents du point de vue géographique et historique sont comparés. Dans le but de cerner la diversité de cette région, on a créé une banque de données. Elle comprend 49 caractéristiques phonologiques et morphologiques permettant la comparaison et l’évaluation historique des dialectes. L’échelle statistique ainsi établie permet d’effectuer des déductions historiques. De plus, deux caractéristiques sont traitées de manière détaillée dans le but de reconstruire des proto-formes à l’aide de la méthode comparative. Les résultats indiquent que l’Arabe parlé avant la diaspora du septième siècle était considérablement plus diversifié que l’histoire de la langue arabe ne le reconnaît traditionnellement.

Zusammenfassung


Author’s address:

Jonathan Owens
University of Maryland
The Center for Advanced Study of Language
PO Box 25
College Park, Maryland, 20742, U.S.A.
E-mail: jowens@casl.umd.edu