Arabic in contact witih other languages

Sarah G. Thomason

University of Michigan

The task assigned to me by the encyclopedia's editors was to read the forty-eight articles that concern Arabic in contact with other languages and to report on their results. The various authors' analyses and conclusions permit a number of generalizations, most of which are unsurprising. More interesting, therefore, are the sometimes quite striking differences among the contact situations and their linguistic and sociolinguistic outcomes. In this article I will survey three sets of topics that recur in the forty-eight articles and discuss briefly some of the implications of these authors' results for general theories of contact-induced language change. Less attention will be paid to the historical, political, and socioeconomic settings of the various contact situations—not because they are unimportant or uninteresting, but because they vary so much: the one thing almost all of them have in common is the crucial role of Islam in the spread of Arabic throughout and beyond the Arabian Peninsula.

The bibliography at the end of the article contains only works that are not included in this encyclopedia; articles that are in the encyclopedia are cited simply by author and article title.

1. *Topics and territories*

By far the most prominent topic in virtually all the articles is the adoption and adaptation of loanwords (§2), primarily from Arabic into other languages but also, in a small number of articles, from other languages into Arabic. This emphasis on loanwords comes as no suprise: with the exception of isolated minority communities of Arabic speakers in Latin America (Capello, 'Latin America') and elsewhere, structural interference is only likely to

have occurred within Arabophone areas; and most of the regions discussed in these forty-eight articles – those on sub-Saharan African contacts, South and Southeast Asian contacts, and even European contacts – are outside Arab lands. The most important subtopics within this area concern the semantic domains of loanwords, the particular word classes that are borrowed, and the phonological and morphological nativization, or lack thereof, of the loanwords. The question of structural interference is explicitly raised much less often, but is especially interesting when it does arise (§3). Next comes a discussion of multilingualism and its manifestations, including language death and pidgins and creoles (§4). The final main section concerns language planning in a broad sense, with special emphases on the goals of teaching Arabic, the choice(s) of writing system(s), and language-purism movements (§5). The sixth and final section is a brief conclusion.

The regions and languages covered in the forty-eight articles can be roughly divided into five groups according to the nature of their contacts with Arabs and/or Arabic. First, the most intimate contacts are (or, in some historical cases, were) with languages spoken within or near Arabic-speaking territory, namely, the Near and Middle East and modern Turkey: South Arabian languages, Aramaic, Modern Hebrew, and Coptic; Berber languages; Persian, Turkish, and nearby Tajik (closely related to Persian) and Tatar (a Turkic language). In these cases – with the possible exception of Coptic, because details of its contacts with Arabic are not well understood – influence between Arabic and the other language(s) has often been mutual and has involved a significant degree of bilingualism. Of the six articles in the set that focus on influence from another language into (regional) Arabic, all but English are in this 'most intimate' group, and the influence of English on Arabic is a relatively recent phenomenon.

In the next three groups, influence has been entirely, or almost entirely, from Arabic into the other language(s), though of course this generalization is not meant to rule out

2

the possibility of other-language influence on local varieties of spoken Arabic, where there were any (but there usually weren't). The second group comprises languages of sub-Saharan Africa, primarily spoken in modern countries lying on or near Arabs' ancient trade routes: Nubian; Somali, Tigrinya, Afar, and other languages of the Horn of Africa; Swahili and other East African languages; Hausa, Kanuri, Songhay, Wolof, Yoruba, Fulfulde, and Bambara. In the third group are languages of Europe, where Arabic once had a dominant presence both in southern Spain and in Sicily and (to a lesser extent) southern Italy, and from those regions influenced other European languages as well, including English. Fourth, and more distantly, Arabic has long been in contact (often indirect contact, via Persian) with languages of South and Southeast Asia, in the Indian subcontinent (in what are now Pakistan, India, and Bangladesh) and in Thailand and Indonesia – especially the latter, with its largely Muslim population. In addition to its spread with Islam, Arabic reached the Indian subcontinent via trade, including trade settlements in Indian Ocean coastal areas. Finally, the fifth group is the isolated case of Latin America, where communities of 19th-century Arabic-speaking immigrants came into intimate contact with Spanish and Brazilian Portuguese; here the pattern was Spanish or Brazilian Portuguese influence on Arabic, rather than vice versa. The Latin American contact situation no doubt resembles isolated communities of Arabic speakers elsewhere, for instance in the United States, but other such contacts have not yet (as far as I know, and as far as the present set of articles is concerned) been studied systematically.

Before beginning the survey, I should mention a methodological issue that arises frequently. Although some authors clearly draw on deep and wide-ranging studies of the various phenomena for their particular locations (for instance Kirchner on Turkish loanwords from Arabic), other authors (for instance Procházka on Turkish loanwords into Arabic and Jakobi on Arabic loanwords into Nubian) warn that no systematic studies have as yet been

carried out on these matters for their locations, so that they can only provide preliminary analyses. Among other things, this problem affects discussions of the numbers and variable phonological nativization of loanwords, and it also hinders attempts to discover structural interference in those (relatively few) cases where one might expect to find some. But the lack of definitive answers in most cases should not be seen as any kind of defect in the authors' analyses; rather, it is an indication that there is still much room for exciting new research on language contacts involving Arabic.

For the most part I will not cite bibliographical sources beyond the articles themselves, because readers can easily find them in the articles. One recent publication, however, deserves special mention: in his 2001 article in *Arabica,* Versteegh conducts a survey quite similar to the present one, drawing on a substantial scholarly literature. He did not, of course, have access to all the papers written for this encyclopedia, but his results are largely confirmed and supported by these authors, and thus by the present article.

2. *Loanwords*

Almost all of the forty-eight articles discuss loanwords, often to the exclusion of other topics. Five articles, in fact, focus explicitly on loanwords, all from other languages into Arabic: Atawneh, 'English loanwords'; Cifoletti, 'Italian loanwords'; Procházka, 'Turkish loanwords'; Zammit, 'South Arabian loanwords'; and Amara, 'Ivrit [Modern Hebrew] loanwords' in (specifically) Palestinian Arabic. One of the most interesting observations in this set of articles is Zammit's mention (citing Jeffery 1938) of South Arabian words as being among the estimated 322 loanwords in the Qur'ān (as Zammit notes, the question of foreign words in the Qur'ān is admittedly a controversial topic). The near-universal focus on loanwords fits the general picture of Arabic in contact with other languages, especially but not only those are spoken relatively far from Arabophone regions. Without the presence of communities of native speakers of Arabic, it is likely that linguistic interference will be

largely or entirely limited to non-basic lexical items. The reason is that intimate contact, including significant degrees of bilingualism, is needed to provide an appropriate social setting for contact-induced language change that affects the receiving language more deeply than the adoption of loanwords; accordingly, with the possible exception of highly formal Arabic-related contexts (such as religion and poetry), deeper influence is found only in Arabophone regions (see §3 for further discussion of this point).

Numerical estimates of Arabic loanwords in other languages are rare, but one figure that is given is startling: Kirchner ('Turkish') says that loanwords from Arabic and Persian once formed more than 80% of the vocabulary of written Ottoman Turkish. (He goes on to emphasize that Ottoman Turkish was never a mixed language, because the grammar and 'verbal core' remained Turkish.) Perry ('Persian') gives a much lower figure for Modern Persian – 8,000 loanwords in a dictionary count and considerably fewer, depending on the genre, in a count of text frequency – but since Persian, like Turkish, underwent a period of attempts to purge its vocabulary of loanwords (see §5), this figure is much lower than it would have been before the 1930s.

Several authors highlight calques (loan translations) in their consideration of loanwords. These are a subtype of lexical borrowing, involving morpheme-by-morpheme translation of words (involving only roots and derivational affixes, not inflection) and/or word-by-word translations of phrases. The most prominent of these discussions are in contact situations in the first group listed above (henceforth Group 1) – the regions in or near Arabophone territory, i.e. the Near and Middle East and modern Turkey. Amara ('Ivrit loanwords') observes that there are many calques from Hebrew in both written and spoken Arabic of the region; and conversely, Geva-Kleinberger ('Ivrit') says that there are many calques from Arabic in Ivrit too, and in this case they come both from 'natural' sources – that is, from people's daily interactions with each other in speech and writing – and from planned calquing, 'a

planned insertion of Arabic loanwords [including calques], especially by numerous innovators beginning with Eliezer Ben-Yehuda and continued by the Hebrew Language Council and, later, by its successor the Academy of the Hebrew Language.' Kirchner ('Turkish') writes that calques, being much harder to identify as of foreign origin, were left as traces after the post-Ottoman Turkish language reform attempted to eliminate all Arabisms. And Procházka ('Turkish loanwords') says that there are many calques in the Arabic dialects spoken by minority groups within Turkey, but that it is often impossible to tell which direction the borrowing went, because of the 'lack of detailed studies of phraseology in both Arabic and Turkish'.

Half the authors comment explicitly on the major semantic domains in which loanwords cluster. These domains are roughly predictable, with one or two surprises, from the nature of the contacts. Words connected with Islamic religious practices and beliefs are prominent in almost all regions with a significant Muslim population, notably in sub-Saharan Africa and South and Southeast Asia; for Hausa, for instance, Abu-Manga estimates that over 50% of the Arabic loanwords concern religion. Strikingly, however, and despite very large numbers of Arabic loanwords in other domains, Persian has not borrowed heavily from Arabic in the domain of religion. Perry ('Persian') accounts for this by the common-sense reasoning that a successful effort to convert people to a new religion will necessarily require that they understand what the missionaries are telling them, so that it makes sense for the missionaries to use the prospective converts' own words in the process. But this explanation won't work for other areas to which Islam spread, as indicated by the extent to which more distant cultures have adopted Arabic religious vocabulary along with Islam. That doesn't necessarily mean that Perry's explanation is invalid for Persian, but it seems more likely that the explanation lies instead, or at least in part, in the prestige of Persian at the time of the conversions, rivaling that of Arabic and far higher than the prestige of most other

6

languages spoken in regions in which Arabic loanwords predominate in the field of religion. It is worth noting in this context that the absence of lexical borrowing is well attested in even quite intense contact situations elsewhere in the world – for instance in Native American languages of the Pacific Northwest of the US and Canada, some of which have borrowed almost no words from English in spite of over 150 years of intimate contact and extreme cultural pressure. Clearly, therefore, cultural factors can and do influence the likelihood of lexical (and other) borrowing.

Other semantic domains are largely or wholly unconnected with religion, except insofar as Arabic-language administration and culture accompanied Islam. The most frequently mentioned domains, obtaining also in Europe and other regions that did not turn to Islam, are trade, science and technology, time, literacy and grammar (in a blow to linguists' self-esteem, Abu-Manga ['Hausa'] lumps grammar with astrology in the category of pseudoscience!), administration, maritime terminology, jurisprudence, food, items connected with daily life (curiously, 'soap' seems to be the item most frequently mentioned), and flora and fauna. Many authors emphasize the very wide semantic range of Arabic loanwords, but overall these domains, especially the first few, reflect the features of Arabic civilization (outside religion) that have been most important internationally: ancient trade routes around the Mediterranean and south to sub-Saharan Africa and the Indian Ocean, and world dominance in science – a dominance still visible in the large number of international words of Arabic origin in mathematics and astronomy. Items of trade (including 'soap'!) and food, as well as other everyday terms, speak to the influence of Arabic culture in near and distant regions.

When Arabic is the recipient language in a contact situation, religion is not (to judge by these articles) a source of loanwords (this is hardly surprising). Instead, everyday items predominate; recent English loanwords in the domains of computers, cars, foods, and clothes provide one obvious example (obvious because English terms in these areas have spread

around the world in recent decades). An interesting example of a more specialized set of loanwords is found in Sudanese Arabic, which has mainly borrowed agricultural terms from Nubian.

Another topic that is covered by many of the authors concerns borrowed word classes. There is quite general agreement that nouns are by far the largest class of loanwords – no surprise there, as this is true all over the world. It is somewhat surprising, however, to find that almost half the articles mention verb borrowing, given the still widespread (though mistaken) view that verbs are rarely borrowed. True, Kirchner ('Turkish') says that finite verbs are 'very seldom' borrowed from Arabic into Turkish; but even this means that at least a few verbs have been borrowed. Verbs have been borrowed into Arabic from (at least) Aramaic, Ivrit, and Turkish, and from Arabic into (at least) Neo-Aramaic, Modern South Arabian languages, Turkish (rarely), Tatar, Persian, Hausa, Yoruba, Fulfulde, Songhay, Swahili, Nubian, Bengali, Telugu, Indonesian, and Malay. Sometimes there are many such loanwords: by one count, there are 75 verbs in Syrian Arabic dialects from Turkish (Procházka, 'Turkish loanwords'). The actual mechanism of borrowing varies. As has been noted for other contact situations around the world, verb borrowing often involves the adoption of a nominalized verb (such as an infinitive or a participle) which is then combined with an auxiliary verb native to the borrowing language; this is found in some of the Arabic contact situations too, for instance in Persian and Bengali. But the language contact literature also has examples of direct borrowing of verbs as verbs, and that too happens in these contact situations – e.g. in Fulfulde and Songhay, where the Arabic imperfect is borrowed, and in the Nubian language Nobiin. By contrast to Nobiin, the closely-related Nubian language Kenzi-Dongolawi borrows verbs by means of an auxiliary -*e(e)* 'say' (Jakobi, 'Nubian'), in a striking instance of different borrowing strategies employed in very similar systems. The overall picture of diverse means and rates of verb borrowing in these contact situations is of considerable sig-

nificance for general investigations of lexical transfer in language contact situations, because it is, to the best of my knowledge, the first collection of parallel studies of the phenomenon in situations involving a single language.

Two other frequently-mentioned borrowed word classes are numerals and discourse markers (both particles and full words). Numerals have been borrowed in (at least) Modern South Arabian languages, where most numerals above 10 are of Arabic origin (Lonnet); Swahili, probably the best-known case, where the numerals for 6, 7, and 9 are of Arabic origin, the numerals for 11-19 are expressed by doublets, both Arabic terms and Bantu terms, and other numerals are Arabic (Baldi); Nubian (Jakobi); Hausa (Abu-Manga); and Javanese (Machali). The borrowing of discourse markers is interesting from a syntactic and pragmatic viewpoint, as such morphemes tend to have structural functions, not (just) lexical ones; no detail is given by these authors on syntactic implications of these borrowings, however. For the rest, some authors say that all word classes are borrowed, including adjectives, conjunctions, prepositions, and various kinds particles, in e.g. Modern South Arabian languages, Hausa, and Nubian, all borrowing from Arabic, and Ivrit loanwords in Arabic. Other languages have borrowed mainly nouns, including some deverbal nominals, and also some adjectives (e.g. Persian [Perry], Indonesian and Malay [Campbell], all from Arabic, and Italian loanwords in Arabic [Cifoletti, 'Italian loanwords']).

One final point on loanword domains: Lonnet discusses a first-person singular suffix borrowed from Arabic into Modern South Arabian. This isolated instance is worth noting because the borrowing of personal pronouns is rather rare in this and most other parts of the world. Kirchner also mentions pronoun borrowing in Turkish from Arabic, but without giving details, so its significance can't be assessed: it is only the borrowing of personal pronouns that is believed to be especially rare. Versteegh has noted that the Arabic pronouns *ane* 'I' and *ente* 'you' are used in Betawi Bahasa Indonesia 'in order to avoid the complicated

9

system of prestige pronouns that exist in many Indonesian languages' (2001:479); pronoun borrowing is actually rather common, for this and other purposes, in Southeast Asia (see Thomason & Everett 2006 for discussion of similar cases).

The phonological, morphological, and syntactic nativization of loanwords is the norm in casual contact situations, where typically only non-basic vocabulary items are borrowed. Only where there is enough knowledge of source-language structure within the borrowing-language speech community is there the possibility of borrowing structure along with words – including the possibility that loanwords will retain some source-language features that are new to the borrowing language. Since the lack of nativization is in fact structural interference, it will be covered primarily in §3 below.

Many of the authors comment on the phonological and, to a lesser extent, the morphological treatment of loanwords in the receiving languages. The main theme is indeed nativization, as one would expect in Arabic contact situations outside Group 1. In most varieties of Hausa, for instance, Arabic sounds foreign to Hausa are replaced by the perceived closest equivalents in the native Hausa inventory, and Arabic loanwords with closed word-final syllables undergo either deletion of the final consonant(s) or epenthesis of a vowel (for instance the suffix -i for masculine nouns) (Abu-Manga, 'Hausa'). The same seems to be true of Bambara (Bouwman), in which (for instance) the Arabic glottal stop is either deleted or replaced by /w, y, h/, and of Fulfulde (Theil), where e.g. Arabic /q/ is replaced by /k/ or /g/ and consonant clusters are either broken up by an epenthetic vowel or simplified by consonant deletion. In most of the languages, emphatic consonants are replaced by non-emphatic counterparts or, when these are lacking, other (fairly) similar phonemes.

Nativization also often occurs when Arabic is the receiving language. Cifoletti ('Italian loanwords') notes that Italian /p/ is replaced in Italian loanwords by Arabic /b/ 'and pronounced voiceless only by cultivated persons'; he also says that initial Italian consonant

10

clusters are broken up by epenthesis, e.g. Cairo Arabic *kire:ma* 'cream' from Italian *crema*. And Procházka ('Turkish loanwords') writes that non-Arabic sounds are usually replaced by native sounds in Turkish loanwords – for instance, /p/ is almost always replaced by /b/.

Although, as we will see in §3, lexical borrowing in these contact situations is sometimes accompanied by phonological interference, at least in Group 1 contact situations, morphological interference is much rarer and is almost entirely confined to a few borrowed plural endings in situations where Arabic is the donor language. That is, morphological nativization is pervasive in these contact situations. Telugu is a typical case. Although, like Arabic, Telugu has a noun-class system that is semantically based partly on biological gender, the two languages differ sharply in their gender categories, and Arabic words are nativized into the Telugu gender system. Moreover, according to Swarajya Lakshmi ('Telugu'), nominals are borrowed as uninflected singular forms and then used with Telugu plural suffixes – in other words, complete nativization. Even in Swahili, where phonological interference from Arabic is significant, Arabic loanwords are generally nativized morphologically (but see §3 for exceptions to this generalization). The other articles make it clear that morphological nativization is complete in most of the other languages.

One final point should be emphasized here: the frequent pattern in which an Arabic noun is borrowed with the Arabic definite article *al-* (or one of its allomorphs) attached is not evidence of Arabic morphological interference in the receiving language. The reason is that the Arabic morpheme and the following nominal are borrowed as an unanalyzable whole. None of the authors of these articles points to any productive use of the Arabic article as a separate morphosyntactic element in the borrowing language; instead, it is a mere phonological part of the noun in the borrowing language, just as an English word like *alcohol*, also originally a borrowing from Arabic consisting of an incorporated Arabic article *al-* plus a nominal stem, is a single English morpheme. Only one author suggests that the Arabic

11

article has morphemic status in the borrowing language: Campbell ('Indonesian/Malay'), citing Verhaar 1984, argues that although *al-* is not productive in Indonesian or Malay, it has 'some degree of morphemic status' – that is, it is identifiable as a morpheme. But he does not say whether nouns occur both with and without *al-*, which would be a requisite for identifying *al-* as a morpheme.

3. *Structural interference*

In spite of the fact that most contacts described in these articles are not intense enough to make extensive structural interference likely, the link between Arabic and one of the world's major religions raises the possibility of relatively minor kinds of structural interference in languages spoken in Muslim areas distant from Arabophone countries. This is especially likely in the writings of highly educated people who have learned Arabic as a second (or third, or...) language. These phenomena will probably be confined mainly to features that enter the language attached to loanwords and are used only with loanwords, but the rather frequent reports of borrowing of conjunctions and discourse markers means that contact-induced syntactic change is also possible (though none of the authors explores this possibility in the present set of articles).

All these are the kinds of features that appear in (for instance) English as a result of borrowing from Latin, which once enjoyed the high level of prestige in Europe that Arabic has in the Muslim world. In addition to the many loanwords that English has adopted from Latin, and the many technical terms coined within English from Latin morphemes, English has a small number of Latin morphosyntactic features, e.g. a handful of Latin plurals such as *alumni* (masc.), *alumnae* (fem.), and *millennia* (compare the singular forms *alumnus, alumna,* and *millennium*). Typically, in such a situation, a language will borrow different forms of the same word separately and then, later on, speakers analyze the forms morphologically. In English, the singular/plural *-us/-i* pattern in particular has become

modestly productive in loanwords, as seen in the innovative (non-borrowed, non-original) plural like *octopi* – where the singular, *octopus*, is ultimately from Greek, not Latin, and was originally a Greek compound *octo-pus*, meaning 'eight feet', which originally had the Greek plural *octopodes*.

Similar kinds of minor structural incursions from Arabic are found in regions outside Arabophone territory that have been influenced by Arabic, for instance broken plurals in Arabic loanwords. Lodhi ('East Africa') observes, for instance, that Swahili has a few of these, though they compete with native Bantu plural formations; an example is *binti* 'daughter', variously pluralized as *mabinti* (with a Bantu plural class prefix) and as *banati* (with an Arabic broken plural formation). Ottoman Turkish, Tajik, and Persian all had broken plurals on some loanwords (see Kirchner, 'Turkish', and Perry, 'Tajik'). Indeed, broken plurals were a target of the 20th-century movement to purge Turkish of foreign elements, and according to Kirchner, the broken plurals that remain in Turkish are lexicalized as singular forms. In Ottoman Turkish, before the language reform, both Arabic **and Persian** loanwords were pluralized with Arabic formations (Kirchner, 'Turkish') – a circumstance which, like the English plural *octopi*, indicates a certain level of productivity of the Arabic plural patterns, even though they remained confined to loanwords. Elsewhere, too, Arabic broken plurals are borrowed intact but lexicalized as singular forms rather than as plurals (see e.g. Perry on Tajik and Campbell on Indonesian/Malay). These lexicalized singular forms of course do not exemplify structural interference, since only the Arabic forms, and not their structure, have been adopted. It is worth noting that although borrowed nouns often have Arabic plural suffixes (and other Arabic features: see §3 below), borrowed verbs are always nativized (Versteegh 2001:479).

Most examples of minor structural interference are reported for the phonology. For instance, Ngom says that some loanwords in Wolof are pronounced with Arabic consonants

not native to Wolof when the speakers have some acquaintance with Arabic – that is, these are learnèd loans. They display such Arabic features as a voiceless uvular stop /q/ and even, apparently, occasional pharyngeal consonants. Cifoletti's report that Italian loanwords in Arabic are generally nativized, but that 'cultivated persons' maintain the distinction between Italian /p/ and /b/ seems to reflect a similar pattern: people who know some Italian do not (always) nativize loanwords completely. And Abu-Manga ('Hausa') differentiates the varieties of Hausa that are spoken in the Sudan from Hausa varieties spoken farther west: in the Sudan, where Hausa speakers have more exposure to Arabic, the speakers sometimes keep emphatic consonants in Arabic loanwords, even to the point of inserting them via hypercorrection where they were not present in the Arabic source word. Both Kirchner ('Turkish') and Wertheim ('Tatar') report that Arabic loanwords often violate the two Turkic languages' vowel harmony rules.

The only hints of more extensive structural interference are found in the articles about contact situations in Group 1, namely, regions in and near Arabophone territory – and also in Swahili, which is a special case. Versteegh (2001:495) reports, for instance, that in Ottoman Turkish one finds 'much more productive' Arabic (and Persian) morphosyntax, including not only the broken plurals mentioned above but also such features as Arabic agreement rules in noun phrases. Citing Prokosch (1980:40), he also notes that such rules were used variably, not consistently, which might reflect a distinction between educated and less educated usage and/or between formal and less formal registers. And Amara ('Ivrit loanwords') observes that some Ivrit loanwords into Palestinian Arabic 'may be used with both Ivrit and Arabic suffixes'; he goes on to say that the choice of suffix often reflects the speaker's level of education, such that highly educated people 'tend to use the Ivrit suffix in most contexts and words'. Occasionally one finds tantalizing comments like Procházka's remark that the dialects of Arabic speakers in Turkey 'are still influenced by Turkish, not only in vocabulary

but also to some degree in morphology and syntax' ('Turkish loanwords') – but without details, which suggests that this is a topic that has not yet been systematically explored.

In the phonology, Arnold ('Neo-Aramaic') identifies chronological layers of borrowing in some Jewish Neo-Aramaic dialects of northern Iraq according to the degree to which Arabic loanwords are nativized: in older borrowings, Arabic phonemes not native to Aramaic have been replaced by native Aramaic phonemes, but in more recent loans, certain Arabic phonemes are retained. Similarly, Jakobi ('Nubian') says that, thanks to Nubian speakers' 'increasing knowledge of and proficiency in speaking Arabic', Arabic loanwords in Nubian sometimes preserve non-native Arabic segments and structures, 'thus enlarging and modifying the original Nubian phonological system'. So, for instance, although emphatics are replaced by non-emphatic consonants and and Arabic consonant clusters are broken up by epenthetic vowels, a voiced alveolar fricative /z/, new to Nubian, is now found in loanwords only, and /l/ and /r/, previously barred from initial position in Nubian, now occur there in loanwords. This picture resembles changing contact situations elsewhere in which an early period of casual contact, with little bilingualism among borrowing-language speakers, saw full phonological nativization of loanwords, while later on, when bilingualism had become widespread among borrowing-language speakers, loanwords were not nativized. A clear example is found in Siberian Yupik (Eskimo), where early Russian loanwords have only native Yupik sounds, but later Russian loanwords preserve (previously) foreign sounds and have thus changed the phonemic inventory of Siberian Yupik (Menovščikov (1969:124-130).

Swahili, with its huge number of Arabic loanwords and its establishment as the major lingua franca of East Africa by Arab traders, lies outside Group 1 regions but nevertheless seems to have undergone more extensive influence from Arabic than have other sub-Saharan African languages within the Arabic sphere of influence. In addition to the broken Arabic plurals on loanwords that were mentioned above, Swahili also has several new phonemes,

confined to Arabic loanwords. According to Baldi ('Swahili'), these are used variably, but especially 'in the speech of Muslim native speakers from the coast, who have had some exposure to Arabic, and for whom pronunciation of these sounds as closely as possible to the Arabic model is a matter of prestige.' Non-Muslim Swahili speakers who have less (or zero) knowledge of Arabic nativize loanwords fully. Register matters: Arabic pronunciation may be heard in 'highly formal' speech, but not in casual speech. The recurrent view that Swahili may have originated as a pidgin or creole, with Arabic as a major component, is rejected by modern scholars (see Nurse 1997). But there is little doubt that the flood of Arabic loanwords has had structural effects, and not only in semantic domains like the numeral system (where most native Bantu numerals have been replaced by Arabic numerals): Swahili is one of very few Bantu languages that have lost phonemic tone distinctions entirely, and the trigger for this development was surely the impact of all those tone-less Arabic loanwords.

Overall, then, the amount of structural interference reported in these articles is nontrivial but also not deep. It is likely that further research, especially on Arabic contacts in Group 1 regions, will reveal considerably more structural interference, not only from Arabic into other languages but also from other languages into Arabic-speaking minority groups.

4. *Multilingualism and its effects*

Not surprisingly, bilingualism/multilingualism in Arabic and other language(s) is concentrated in Group 1 regions that have a major Arab presence. Lonnet, for instance, writes that almost all speakers of Modern South Arabian languages (except in Soqotra) speak their own language, one or two other Modern South Arabian languages, and also Arabic; and Hausa-speaking communities in Sudan, Libya, and Saudi Arabia tend to be bilingual in Arabic and Hausa (Abu-Manga, 'Hausa'). The picture is often less clear for past eras – the question of when, and even whether, bilingualism obtained in Andalusia is highly controversial, for instance (Zwartjes, 'Andalus') – but some authors can confidently discuss past bilingual-

16

ism in, for instance, Aramaic-speaking regions (Retsö, 'Aramaic/Syriac'), Sicily (Metcalfe, 'Sicily'), Persia (Perry, 'Persian'), and certain Indian towns whose residents were bilingual in Sindhi and Arabic (Qutbuddin, 'India', citing Yusuf 1967:56). Capello's description of the ups and downs of bilingualism in Arabic and Portuguese (Brazil) and in Arabic and Spanish (especially Argentina) is particularly interesting, covering such factors as the speed of assimilation of Arabic-speaking immigrants and the wider community's negative or positive reactions to Islam. His article ('Latin America') also highlights, by contrast, the extent to which elucidation of the political and social conditions of past contact situations is hampered by incomplete information: even with all the documentation available on the former use of Arabic in Persia, Turkey, Andalusia, Sicily, and elsewhere, it remains impossible to carry out the kinds of fine-grained analyses that one can conduct on a live contact situation, and generally impossible to speak with confidence about the linguistic repertoires of illiterate folk, as opposed to literate, educated higher strata of society.

A few authors comment on the recent decline in the use of Arabic in public life in regions outside Group 1, for instance on the Tanzanian mainland, where one no longer sees shop signs and other signage in Arabic in the towns (Lodhi, 'East Africa'). But the instances of contact situations within Group 1 regions in which Arabic is replacing other languages – that is, in which language death is occurring – are more striking: Some South Arabian languages are being replaced by Arabic (Lonnet, 'Modern South Arabian'); and several Nubian languages have already vanished as a result of Arabization (Jakobi, 'Nubian').

A parallel topic, one that concerns the most dramatic things that happened to Arabic as it spread rather than what happened to the languages with which it came into contact, has to do with Arabic-lexifier pidgins and creoles. The topic is addressed in this set of papers by Owens' article 'Creoles', and it seems to provide at least some of the background for Simeone-Senelle's article on Arabic as a lingua franca in the Horn of Africa; but it is also

a topic that has received a significant amount of attention within Arabic studies over the past twenty years. The first major work to focus on pidgins and creoles in this domain was Versteegh 1984, an investigation of the possibility of pidginization as a phenomenon accompanying the spread of Islam to what are now Arabophone territories; two years later, Prokosch (1986) surveyed Arabic-based pidgins and creoles in Africa, where almost all the ones reported in the literature are located. Owens himself has written extensively on this subject, including at least two general survey articles in addition to the present one (Owens 1997, 2001). Proposals about Arabic-lexifier pidgins spoken early in the history of the spread of Arabic and Islam have been based on fragmentary documentation from as early as the 11th century CE (Thomason & Elgibali 1986, reporting on a passage in al-Bakrī) and on inferences about the kinds of contact situations that likely arose as Arabic spread (e.g. Versteegh 1984). I will not address this topic in detail here because for the most part, though it is obviously related to the general subject of Arabic contacts with other languages, it does not lend itself to direct comparisons with other kinds of contact situations – namely, those whose social contexts and linguistic results were less drastic.

5. *Language planning*

A final recurring set of topics in the articles surveyed here falls under the general heading of language planning. Many of the authors discuss the teaching of Arabic in various regions, the choice(s) of a writing system for other languages in regions influenced by Arabic, and language reform movements designed to purge various languages of Arabic loanwords. These are all huge topics, of course, and space limitations make it impossible to do justice to them, especially as policies and practices have changed over time in a number of regions, sometimes more than once. This section, therefore, is merely a sketch of the issues that arise in this area.

The teaching of Arabic outside Arabophone regions has often, as one might expect, aimed

at a reasonable level of competence in written or spoken Arabic, or both. But sometimes the goal is instead to teach the rudiments of the alphabet and the recitation of prayers, nothing more; comprehension of the Arabic prayers is neither required nor expected. There are various religious and political motivations for the latter practice, and it is tolerably widespread in Muslim regions; it is mentioned especially in the articles on South Asia (Qutbuddin, 'India'; Rahman, 'Pakistan'; and Riaz, 'Bangladesh'). Rahman writes that, in the 18th century, 'The Arabic script...remained part of the traditional course of studies of a Muslim gentleman and even women, while denied literacy in other languages, were taught how to read the Qur'ān without understanding. The pupils merely learnt to recognize the Arabic alphabet before going on to studying Persian, a marker of elitist identity as well [as] the language of upward social mobility.' Various other political and social motives have driven Arabic teaching in these and other regions, and the articles in this set offer rich material for the study of the issues.

A related topic is the choice of one or more writing systems for languages in contact with, and under the influence of, Arabic and (usually) Islam. This matter is at least mentioned in almost a third of the articles, an indication of its central importance in language planning: it is a topic with profound political implications. Modern Persian, for instance, is written with a version of the Arabic alphabet, but Middle Persian was not – it was written in an Aramaic script (Perry, 'Persian'). The change to the Arabic alphabet came with Islam by the mid-9th century CE. Ottoman Turkish was written with Arabic letters, but Modern Turkish is written in the Latin alphabet; the change was made as part of the language reform undertaken by a secular government looking to the West for its future. The history of Tatar writing has been unusually complex (Wertheim, 'Tatar'): in the mid-19th century, Tatar was written in the Arabic alphabet; in 1927, the Arabic alphabet was abandoned in favor of Latin letters; and in 1938, in accordance with Soviet policies regarding minority

19

languages, a Cyrillic alphabet replaced the Latin one.

In Malta (Brincat, 'Malta'), where the spoken Arabic variety lost its cultural connection to Classical Arabic very early, Maltese Arabic was never written in the Arabic alphabet; even the earliest texts, starting in ca. 1470 CE, were written in Latin letters. In Sub-Saharan Africa, the first writing systems were often Arabic (Baldi, 'Swahili'; Ngom, 'Wolof'; Theil, 'Fulfulde'). Under European colonial rule, literate Swahili speakers generally knew only the Arabic alphabet, but the colonial government in Zanzibar persisted in publishing Swahili in the Latin alphabet (Lodhi, 'East Africa') – a practice that guaranteed that almost no Swahili speakers would be able to read it. In South and Southeast Asia, Urdu, an official language of Pakistan that is also widely spoken in India, is written in Arabic script, while Hindi, an official language of India, is written in a Sanskrit-derived script; and yet Urdu and Hindi are so closely related as to be arguably dialects of the same language. Farther east, a comparable split is found in Indonesia: until the end of the 19th century, Malay was (and sometimes still is) written in a version of the Arabic alphabet, while certain other Indonesian languages – notably Javanese – still use a Sanskrit-derived writing system (Steenbrink, 'Indonesia'; Campbell, 'Indonesian/Malay').

This brief survey of alphabet choices necessarily omits a great many useful and important details in the various authors' discussions. It does, however, provide an indication of the spread of Arabic writing in addition to, and sometimes partly independently of, the spread of the Arabic language itself. A very recent development in Morocco underscores the cultural and political dimensions of such a choice. With the newly-established policy of introducing Berber into some schools came the necessity for standardizing Berber (specifically Tamazight Berber), including establishing an official writing system. The planners rejected both the Latin alphabet, formerly a symbol of the French colonial government, and the Arabic alphabet, the writing system of the country's official language, Arabic; instead, they chose

to revive the ancient Tifinagh writing system because it is unique to Berber (Hamid Ouali, personal communication, 2005).

The move to establish Berber in Moroccan schools as a language independent of Arabic (and French) is just one of many instances in which speech communities have tried to free themselves of the cultural weight of Arabic (although in most cases they continue to revere Arabic as the language of the Qur'ān). Perry ('Persian') and Kirchner ('Turkish') describe vigorous debates and policies on this issue in Persia/Iran and Turkey, respectively. The use of Persian itself in scholarly writing was controversial in Persia a thousand years ago; that changed in the 13th century, when Arabic ceased to be spoken in Persia, but it was not until the 1930s and 1940s, in Iran, that language reformers attempted to replace Arabic loanwords with native Persian words. They were less successful (or less radically inclined) than the language reformers in Turkey. There, during the transition from the Ottoman Empire to the modern Turkish state in the 1920s, the politically dominant radical purists wanted to eliminate all foreign elements – especially those from Arabic and Persian – from Turkish, and as a result of their efforts the percentage of Arabic (and Persian) loanwords in Standard Turkish was sharply reduced. Similarly, Tatar lost most of its Arabic and Persian loanwords as a result of Soviet policies, which led to a wholesale replacement of those words – not by native words as in Iran and Turkey, but by Russian words; nowadays, half the words in a Standard Tatar-Russian bilingual dictionary are of Russian origin, though since perestroika some Arabic words have again been appearing in written Tatar (Wertheim, 'Tatar'). In old Bengal, an influential 18th-century grammarian, Halhed, 'considered foreign elements pollutants in the "pure" Bengalese'; but the 19th century saw an influx of Perso-Arabic borrowings into Bengali in reaction to the British/Hindu Sanskritization of the language (Wilce, 'Bengali'). Differences of opinion about the cultural meaning of loanwords remain, however.

A very different form of linguistic purism was inflicted on, and resisted by, the Maltese speech community (Brincat, 'Malta'). The British, while trying to replace Italian with English on the island, also promoted Maltese; but their efforts along these lines involved trying to make Maltese a 'purer' language by bringing it closer to Classical Arabic, an approach that met with strong negative reactions among the community's educated elite.

6. *Conclusion*

Anyone who reads this set of forty-eight articles will get a comprehensive picture of Arabic in contact with other languages, with all the rich variety of linguistic, social, and political settings. The picture is not complete, of course: most of the issues discussed above still need further systematic investigation, including both case studies of individual contact situations and comparative studies of partly similar contact situations. Comparative information is already available in these articles (and in the literature they cite) on some topics, for instance strategies of verb borrowing, patterns of phonological nativization of loanwords, and motivations for choosing writing systems. The study of other topics is barely touched on in these articles, perhaps most notably the issue of morphosyntactic structural interference; if its absence in this encyclopedia reflects the state of research on the topic, this is clearly a growth area for Arabic scholarship.

Bibliographical references

Jeffery, Arthur. 1938. *The foreign vocabulary of the Qur'ān.* Baroda: Oriental Institute.

Menovščikov, G.A. 1969. O nekotoryx social'nyx aspektax èvoljucii jazyka' [On some social aspects of the evolution of language]. *Voprosy social'noj lingvistiki*, 110-134. Leningrad: Nauka.

Nurse, Derek. 1997. Prior pidginization and creolization in Swahili? *Contact languages: a wider perspective,* ed. by Sarah G. Thomason, 271-294. Amsterdam: John Benjamins.

Owens, Jonathan. 1997. Arabic-based pidgins and creoles. *Contact languages: a wider perspective,* ed. by Sarah G. Thomason, 125-172. Amsterdam: John Benjamins.

Owens, Jonathan. 2001. Creole Arabic: the orphan of all orphans. *Anthropological Linguistics* 43:348-378.

Prokosch, Erich. 1980. *Studien zur Grammatik des Osmanisch-Türkischen unter besonderer Berücksichtigung des Vulgärosmanisch-Türkischen.* Freiburg: K. Schwarz.

Prokosch, Erich. 1986. *Arabische Kontaktsprachen (Pidgin- und Kreolsprachen) in Afrika.* Graz: Institut für Sprachwissenschaft, Universität Graz.

Thomason, Sarah G., and Alaa Elgibali. 1986. Before the Lingua Franca: pidginized Arabic in the eleventh century A.D. *Lingua* 68:317-49.

Thomason, Sarah G., and Daniel Everett. 2006. Pronoun borrowing. *Berkeley Linguistics Society* 27 (in press).

Verhaar, John W.M. 1984. Affixation in contemporary Indonesian. *Towards a description of contemporary Indonesian: preliminary studies,*, I, ed. by Bambang Kaswanti Purwo, 1-26. Jakarta: Universitas Atma Jaya.

Versteegh, Kees. 1984. *Pidginization and creolization: the case of Arabic.* Amsterdam: Benjamins.

Versteegh, Kees. 2001. Linguistic contacts between Arabic and other languages. *Arabica* 48:470-508.

Yusuf, S.M. 1967. Arabic language and literature in the Indo-Pakistan subcontinent. *Iqbal* 16/1:56-66.