16

Orthography development

Friederike Lüpke

16.1 Introduction

> Vernacular literacy involves much more than merely devising the optimal orthography for a given language as many linguists would have us believe. (Mühlhäusler 1990: 205)

Many endangered languages are not written; therefore, researchers and speech communities often wish for their GRAPHIZATION (Fishman 1974). The existence of a written code is seen as an essential prerequisite for many activities in favour of their maintenance and revitalization, such as dictionary writing (see Mosel, Chapter 17), curriculum development and the design of language-teaching courses (see Coronel-Molina and McCarty, Chapter 18).

Graphization or orthography development is a complex task which requires a careful assessment of issues going beyond purely linguistic decisions. The successful creation of an orthography involves the consideration of historical, religious, cultural, identity-related and practical factors in addition to linguistic ones. Although writing in the mother tongue is recognized as an important linguistic right, literacy can only be successful if there are adequate and varied materials available for reading (and instruction). This means that the potential role and scope

Many people have contributed to shape my thoughts on this issue. First and foremost, I am indebted to the Jalonke speech community in Saare Kindia (Guinea) for sparking my initial interest in orthography development. The participants and audiences of the ELAP workshop on training and capacity building for endangered-language communities in February 2004, on endangered-languages and literacy in December 2005, and of the ESF exploratory workshop on contact between Mande and Atlantic languages in September 2008, all held at SOAS, London, got me further engaged with the subject. Finally, I would like to thank the following colleagues at SOAS for sharing their experiences with me: Peter Budd, Gerardo De Caro, Stuart McGill, Julia Sallabank and Lameen Souag. The responsibility for any misrepresentation of their views lies exclusively with me.

of literacy (as a social practice rather than a technical skill) in an endangered language needs to be evaluated prior to orthography development, and that graphization has to be embedded with care into the larger task of 'corpus planning' (Kloss 1968, Sallabank, chapter 14).

Endangered languages are usually spoken in multilingual environments, and in most instances at least one contact language[1] already exists in a written form and is used for formal contexts of writing. It is therefore important to identify an ecological niche for writing in the endangered language, that is, registers and contexts which are predisposed for writing in it instead of in a contact language. If and when such a context has been found, a writing system and script need to be selected.

A writing system is the abstract underlying type (for instance LOGOGRAPHIC, SYLLABIC, ALPHABETIC, etc.) of which scripts (i.e. Arabic, Latin, Devanagari, etc.) are instances. Scripts are not identical with orthographies/spellings, the standardized versions of scripts for specific languages or varieties thereof (e.g. American versus British spelling). Many twentieth-century REFERENCE ALPHABETS (e.g. the Bamako 1966 and Niamey 1978 alphabets for African languages), are based, in the colonial spirit, on the Latin script, and can ultimately be linked to missionary societies who commissioned the first unified reference alphabet (Lepsius 1863).

It is often assumed that the writing systems of modern orthographies will be of the alphabetic type, but other writing systems persist and need to be taken into account. In addition to preferring alphabetic writing systems, within this type many linguists may lean towards the Latin script because of its closeness to the Latin-based International Phonetic Alphabet (IPA) in which they often produce phonetic and phonological transcriptions. However, in many areas of the world, alternative scripts exist, and in these areas script choice requires conscious and informed decisions. Once this hurdle is overcome, a number of analytical and practical issues need to be addressed.

The written use of a language presupposes its standardization, which is often seen as concomitant with writing. Depending on the internal diversity of the endangered language and the attitudes of speakers to its different varieties, there are several possibilities: creating a KOINÉ variety, an underspecified orthography, or promoting one variety to standard by basing the orthography on it. These choices can have far-reaching consequences on linguistic diversity, the ecological equilibrium of varieties involved, the acceleration of cultural change and loss of the phatic values of the vernacular, as Bielenberg (1999), Mühlhäusler (1990) and Sallabank (2002) warn.

Since Pike (1947) it has become is customary to regard orthographies as 'optimal' if they adhere to the often invoked 'phonemic principle' according to which a one-to-one relationship between phonemes and GRAPHEMES is ideal. Yet, scholars of writing do not cease to stress the differences between orthographies and phonetic or phonological transcriptions (Coulmas 2003, Venezky 2004), so thought must be given to the number and shape of graphemes and their relationship to the phonemes and phones of the language as well as to criteria determining word boundaries. Finally, reflections on how to facilitate the creation and sustainability of a written environment are in order if the orthography is meant to have a lasting impact.

Section 16.2 starts by exploring the ecology of writing in endangered-language communities discussing spoken and written repertoires (16.2.1), DIGRAPHIA (16.2.2), EXOGRAPHIA (16.2.3) and the significance of global narratives of writing and education in this context (16.2.4). Section 16.3 identifies the main issues at hand when choosing a writing system or script, touching on the relations of script with identity and religion (16.3.1), investigating whether there is a natural proclivity of certain scripts to be used for particular languages (16.3.2), and concludes with a discussion of practical matters associated with script choice (16.3.3). The non-linguistic, linguistic and practical questions surrounding orthography development are examined in Section 16.4. Just like scripts, orthographies reflect traditions and identities, and this function is discussed in Section 16.4.1. Section 16.4.2 offers general design considerations for non-logographic orthographies[2] stemming from psycholinguistic research on reading and writing. Section 16.4.3 is dedicated to the practical consequences of orthographic choices. The conclusion reflects to what extent universal discourses on the languages used for writing and education and their roles reflect the linguistic, cultural and socioeconomic realities of endangered-language communities in different endangerment situations and what realistic expectations for the role of writing and the scope of literacy in endangered languages might be.

16.2 The ecology of writing in multilingual endangered language communities

> The bilingual is *not* the sum of two complete or incomplete monolinguals; rather, he or she has a unique and specific configuration … The bilingual uses the two languages – separately or together – for different purposes, in different domains of life, with different people. (Grosjean 2008: 13–14)

This section first presents the different ways in which spoken and written modalities interact in, typically multilingual, endangered language communities by giving an overview of repertoires and functions often associated with endangered and contact languages in the two MODALITIES. Writing traditions using one script versus multiple scripts/orthographies are discussed and contrasted with situations characterized by the total absence of writing. The notion of DIGRAPHIA, often used to characterize multigraphic practices, is introduced, and its usefulness scrutinized in

16.2.2. Exographia, or the absence of vernacular writing, is discussed in 16.2.3. The concept of ecology of writing, inspired by the concept of ecology of language (Mufwene 2001), which does not see contact languages globally in competition with each other but rather understands them as competing for functions, is then compared with global narratives of writing and education with more essentialist assumptions on (oral and written) language use in Section 16.2.4.

16.2.1 Spoken and written repertoires in multilingual speech communities

It is rarely the case in multilingual speech communities, even those using major languages, that their members have identical repertoires in all languages. This observation holds at the level of the oral modality and even more so for the written modality. In contrast to spoken language, writing is not acquired by exposure over a long period of time at a young age, but by more regulated apprenticeship, generally associated with some form of schooling, and requiring technology (stylus, pen, paper, parchment, slate, word processor, etc.) Since writing is more 'costly' than speaking, it is a safe assumption that there will be even less overlap in written repertoires than in spoken ones, i.e. it will be more improbable to find two written languages in a speech community being used for the same functions and contexts than to find overlap in spoken repertoires.

An example from my own experience, the endangered Mande language Jalonke, spoken in Guinea, West Africa (Lüpke 2004, 2005), may serve to illustrate this point. In the local speech community, Jalonke is confined to the oral sphere and has mainly the status of a home language. In all public contexts, the contact language Fula is spoken. Written communication regarding personal and religious matters, and book-keeping at the village level, takes place in Fula, in an Arabic-based script. Written interaction with the authorities and official documents is in French, the official language of Guinea. Each of the languages thus occupies its own ecological niche with very specific functions for the spoken and written modes. If one wished to develop a written code for Jalonke, a careful consideration of its purpose would be required.

Similarly, speakers of the endangered Austronesian language Touo, spoken in the Solomon Islands, employ a variety of Touo and Solomon Island Pijin (Terrill and Dunn 2003). The language learned at school used to be Roviana, another contact language, when it was taught in Methodist schools, until Roviana was replaced by English in this context. Depending on their Christian creed, Touo speakers now write different contact languages in informal contexts: community members who are Seventh Day Adventists are more exposed to writing in the contact language Ughele used by missionaries of this creed, whereas members of churches which descended from the Methodist mission are still exposed to written Roviana. Since the orthographies for the two languages follow different design principles, the delicate problem of avoiding religiously motivated digraphia for Touo poses itself to orthography developers (see 16.2.2 below).

16.2.2 Digraphia

Digraphia is a concept with two different interpretations. For some (DeFrancis 1984, Humery forthcoming, Zima 1974) it is used by analogy with the term diglossia, which according to Ferguson (1959) describes a situation in which two or more language varieties which are used by the same community but are employed in separate contexts and functions, usually considered to be in a hierarchical relationship and hence labelled H (for 'high') and L (for 'low'). On this reading, digraphia only denotes multigraphic writing traditions in contexts where one of the traditions is the dominant one, either synchronically or diachronically. Others (e.g. Coulmas 2001, Grivelet 2003: 231) disregard this interpretation and understand digraphia to 'simply' mean 'the use of two different scripts, writing systems or orthographies for the same language'. I consider it useful to reserve the term digraphia for hierarchical separated functional relationships between written codes, and use the more neutral terms bigraphia or multigraphia (henceforth used interchangeably), coined following the example of bilingualism and multilingualism, for the simple coexistence of two or more written codes for a language or variety (see Fishman, 1967 for an analogous proposal of multilingualism).

Both digraphia and multigraphia are common for languages which have, for a variety of reasons, come into contact with more than one written code, and there are many textbook examples available for larger languages (e.g. Hindi and Urdu, Serbian and Croatian, Chinese characters versus Pinyin[3], etc.). One would hope that digraphia and multigraphia would not be an issue for minority and endangered languages, since their existence increases the complexity of creating and maintaining a written ecology for these languages even more, but unfortunately this is not the case. Touo, mentioned above, is a case in point. Terrill and Dunn (2003) were facing the problem that, depending on their religious orientation, speakers of this language (which has only approximately 1,800 speakers), favoured either a 'Seventh Day Adventist' orthography based on the contact language Ughele, or one of Methodist provenance based on the contact language Roviana, and were not prepared to accept a compromise. For Touo, this unfortunate situation stems from non-coordinated orthography creation by missionaries with different affiliations.

An additional example of multigraphia from my own research concerns the endangered Atlantic language Baïnouk, spoken in Senegal

Table 16.1. *Differences between NTM alphabet and national alphabet for Baïnouk with IPA correspondences*

NTM grapheme	National alphabet grapheme	Corresponding IPA symbol
<a>	<a>	[a], [ɑ], [ɐ]
<e>	<e>	[ɛ]
<i>	<i>	[ɪ]
<o>	<o>	[ɔ]
<u>	<u>	[ʊ]
<á>	<ë>	[ə]
<é>	<é>	[e]
<í>		[i]
<ó>	<ó>	[o]
<ú>		[u]

(West Africa). Here, missionaries of the New Tribes Mission (NTM) created an alphabet for one variety of the language, without taking orthographical conventions existing at the national or regional level into account. When the Baïnouk speech community applied for 'codification', of the languagee – that is, its recognition as a national language with the right to be used in the public sphere – the existing NTM alphabet needed to be adapted to the standard (see Table 16.1 for correspondences). In the NTM alphabet, closed vowels have an acute accent above the vowel grapheme. However, <á>[4] is used by the NTM alphabet to write the schwa sound [ə][5], and hence a closed [ɐ] is not written <á>, breaking the logic of notating closed vowels with an acute accent for the other vowels. In the national alphabet, <ë> stands for schwa, and degree of aperture is only distinguished for the front mid vowel pair and the back mid vowel pair. In view of these inconsistencies, all existing literacy materials for Baïnouk became obsolete overnight.

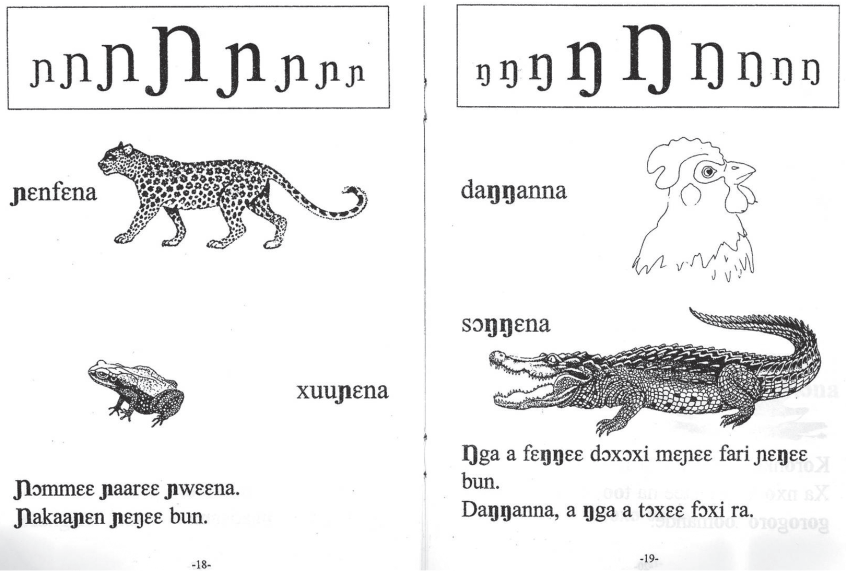
Dominant personalities and/or cultural and religious institutions can have a huge impact on how an endangered language is written, and in many cases it will be difficult, if not impossible, to reverse resulting multigraphia once it is established. The continuation of this variability (rather than pressing for standardization) can be adopted by language activists, as in the case of the endangered language Guernesiais or Guernsey French, a Norman language spoken on Guernsey, one of the Channel Islands. Sallabank (2002: 241) reports the following note from the *Bulletin of L'Assembllaie d'Guernesiais*: 'Notaai s'y vous plait: L'Epellage des les articles du Bulletin a etaai lesi a la discretion des contribuables. [Please note: spelling in the articles of the Bulletin has been left to the discretion of the contributors.]' Another newspaper, the *Globe*, adopts a similar stance to variation, and Sallabank (2002: 231) lists some examples; for instance the Guernesiais form for 'young' written as <jeuaune>, <jeonne> and <jonne>.

Multigraphic practices can come into existence when endangered-language communities are dispersed over territories belonging to different countries with different national script traditions and standards. The speakers of the Indo-Iranian language Taleshi, for instance, are found in northern Iran and Azerbaijan, a former Soviet Republic. This endangered language has been written using Arabic, a modified Cyrillic alphabet, and a number of modified Latin alphabets: the Azeri alphabet introduced in 2001 (which replaced the Cyrillic alphabet in Azerbaijan), and modified IPA-based scripts (Gerardo De Caro, p.c.). Linguists and activists aiming at long-lasting usability of their orthographies and literacy materials are therefore advized to survey existing writing traditions in the endangered language and surrounding languages, existing conventions and recommendations at higher levels, and to consult members of the endangered-language speech communities in order to avoid digraphia or multigraphia or to minimize its divisive effects by, for instance, producing multigraphic materials or creating transliteration guidelines or computer programmes that will map between scripts.

16.2.3 Exographia

I use the term EXOGRAPHIA to designate writing which takes place exclusively in another language. Exographia is very widespread in endangered and minority languages for which no written variety is available at all. It is often the case that an official language (often an ex-colonial one) occupies formal writing contexts and a regional lingua franca is used for writing in semiformal and informal contexts such as adult literacy campaigns, the writing of personal letters, etc. Exographic writing traditions are often overlooked or marginalized (see 16.2.4 below), but it is always worthwhile to conduct a detailed study on functions and uses of writing in other languages prior to embarking on orthography development for an endangered language. If no ecological niche for writing can be found for the endangered language, exographia may be its fate, and it is disputable whether this is cause for concern or not. Endangered languages are often used in small-scale rural communities whose members see each other on a daily basis. If they already have another language at their disposal for writing and if this language is larger and consequently more able to offer a satisfactory written environment, then there may be no need for writing in the endangered language, unless the resources to support long-term DOMAIN EXPANSION are available, e.g. as part of a revitalization programme (see Hinton, Chapter 15).

Many fieldworkers report that finding appropriate contexts for writing is the biggest obstacle they encounter. Often, literacy materials produced in the endangered language are warmly welcomed because of the prestige they lend to the language, but they have little or no practical use because established exographic traditions pre-empt the introduction of



ɲ ɲ Ɲ Ɲ ɲ ɲ ɲ ɲ

ɲenfɛna

xuuɲɛna

Ɲɔmmɛɛ ɲaarɛɛ ɲwɛɛna.
Ɲakaaɲen ɲɛŋɛɛ bun.

-18-

ŋ ŋ ŋ ŋ Ŋ Ŋ Ŋ Ŋ Ŋ

daŋŋanna

sɔŋŋɛna

Ŋga a fɛŋŋɛɛ dɔxɔxi mɛɲɛɛ fari ɲɛŋɛɛ bun.
Daŋŋanna, a ŋga a tɔxɛɛ fɔxi ra.

-19-

Figure 16.1. *Sample pages of the Jalonke primer using a Latin-based orthography (Lüpke* et al.*, 2000)*

ENDOGRAPHIC ones for the same functions.[6] I had this experience in my own research on Jalonke, when I developed a primer using a Latin-based orthography (see Figure 16.1). Although the primer was in high demand and even speakers of the dominant contact language Fula queued for their copy, nobody except my two main language consultants ever wrote in this orthography. The established literacy practice in this endangered language community is to write in Fula, using an Arabic-based script commonly used in the area and carrying strong positive connotations such as links to Q'uranic scholarship.

16.2.4 Writing endangered languages and global narratives of writing and education

The stance towards exographia taken in 16.2.3 above is in stark contrast with discourses of language rights that promote endographia or writing in the 'mother tongue'. Advocates of linguistic human rights (e.g. Skuttnab-Kangas and Phillipson 1995) stress the cognitive and psychological advantages of learning to read and write in and through the mother tongue as opposed to a foreign language, and indigenous literacy is seen as an important factor for language maintenance (Crystal 2000, Fishman 1991). At the same time, numerous political, practical, financial and communicative obstacles to the implementation of mother-tongue education have been identified, especially in endangered and minority language communities (Fishman 1995, 2001b; Romaine 2006b; Spolsky 2004).

Side-stepping feasibility issues, I would like to pause and consider the very notion of 'mother tongue' and 'writing in the mother tongue' and its universal applicability. There are instances where it is impossible to identify the mother tongue of a multilingual individual or the first language of an endangered language community unequivocally.[7] For the African context, for instance, current recommendations for language teaching, for instance from the UNESCO Institute of Education, avoid the term 'mother tongue' altogether and stress instead the advantages of using a familiar language, which in most cases will be an African contact language, as the medium of instruction. This development takes into account the difficulties of unequivocally identifying a mother tongue in contexts of extensive multilingualism (see Blench 1998 and McLaughlin and Sall 2001 for African cases, and Evans 2001 for similar observations on Australia).

Where exographic traditions exist, it may be useful to distinguish two radically different types:

1. a situation where a written majority language with close cultural and/or linguistic affiliations is already present in the multilingual repertoire;
2. exographic practices using an official (often ex-colonial language) that is not part of the everyday repertoire of the endangered language community (for instance the official languages in most countries of sub-Saharan Africa, which have to be acquired in spoken and written modes while at the same time serving as the medium of instruction).

Another widespread but problematic belief is the necessity for a language to be written in order to be a fully fledged language. The existence of a written form lends almost mythical qualities to a language. This language ideology, which Blommaert (2004) calls 'graphocentrism', means that revitalization and maintenance campaigns for minority and endangered languages often focus on the introduction of writing (see 16.2.3 above). While the cognitive and socioeconomic benefits of literacy are undisputed, it is an open question whether this literacy needs to be endographic in all cases, or whether certain exographic approaches may have equally positive effects.

The development of an orthography is often seen as an essential component of language documentation. Seifart (2006: 275) argues that:

> [m]uch of the success of a language documentation depends on casting these records in an orthography that appeals to the speech community. As a matter of fact, if it is accepted that the documentation has to be accessible to the speech community, the development and implementation of a practical orthography in the speech community is an absolutely necessary task in an early phase of a documentation project.

While I absolutely agree with the tenet of making documentation accessible to the endangered language speakers, I would like to propose that developing an orthography is no longer necessarily the most suitable way to achieve this goal. In the past, when written documents were the only type of documentation produced by linguists, the accessibility of the language indeed depended on an accurate rendition of its pronunciation, although it is impossible for an orthography to entirely achieve this (see also Coulmas 2003: 26–35 on the differences between transcription and orthography). Even the most faithful transcriptions are limited in terms of what they represent (e.g. the segmental phonology of consonants and vowels, but ignoring prosodic features of spoken language), and so it is doubtful that spoken language can be rendered in all its facets by any transcription system in use today. Modern technology, however, has enabled language documentation to make audio-and video records accessible to speakers of endangered languages without having to resort to a written representation. Fluent speakers rely much less on phonological information in reading than language learners (among them outside linguists). Semi-speakers and rememberers (see Grinevald and Bert, Chapter 3, for these terms) can learn the language based on audio- and video-records, their transcriptions and annotations. The presentation of oral genres in oral formats (annotations notwithstanding) also preserves their distinct nature in terms of genre, variation, phatic value, etc. and allows the delicate issue to be side-stepped of how to render communicative events of predominantly oral languages in written form, or what new written genres to create.

A written form for their languages features among the strongest wishes of many endangered language communities. However, these positive attitudes towards literacy are not necessarily matched by actual literacy practices. In my research on Baïnouk, 97% of the speech community reported seeing literacy in their language as very positive; however, only 22% attended the literacy classes offered by NTM missionaries, which have now stopped. There are numerous accounts of unsuccessful literacy programmes, especially in developing countries, signalling the huge challenges to be overcome, and these campaigns focus on majority languages for the most part (see Dumestre 1994, 1997 and Prah 2001 for some African observations, and Elwert 2001 and Triebel 2001 for general discussion).

Unless there is a real need and willingness to introduce endangered language literacy in the community, and unless this is backed up by adequate resources, I therefore consider a consistent and documented transcription sufficient and would recommend disseminating audio- and video-records as widely as possible by copying, distributing or broadcasting them, instead of trying to introduce endographia against all odds.

Transcriptions and dictionaries can supply evidence that the language can be written, contrary to popular beliefs, and make a number of emblematic documents available. A successful orthography, however, requires a much larger investment, including:

- selection of a writing system and set of graphemes
- establishment of rules specifying the relationship between sounds (PHONES and PHONEMES) and graphemes
- determination of rules specifying word boundaries and punctuation
- production of a dictionary listing spellings and materials for learning and later independent reading, etc.

16.3 Choosing a script

> The place of writing systems in the study of language planning and language policies is often seen as secondary. The various questions related to writing, such as the choice of writing systems, the type of orthography, etc., are often understood as being obvious, based on two main assumptions: first, that the Latin script is the most suitable to form the base of a new writing system; and second, that a writing system should be phonemic. However, these answers are mainly based on linguistic observations, without much concern for the place and role of a writing system in society. (Grivelet 2001: 1)

This section identifies the main factors in deciding on a script. It starts with investigating the relationship between script and religion and other aspects of historically grown identity that need to be taken into account. Section 16.3.2 discusses a myth circulating among linguists, educational practitioners and speech communities that some scripts are better suited for the writing of particular languages than others. This section illustrates how symbols can be adapted, their inventory extended and the type of writing system matched to the structure of a new language. Section 16.3.3 addresses a number of practical questions related to script choice, such as its consequences for the use of technology (and vice versa) and the production of written materials.

16.3.1 Script, religion and identity

The famous maxim ‘alphabet follows religion’ (Diringer and Regensburger 1968) stems from the observation that the spread of writing systems is largely coextensive with that of the world’s major religions. Religion is a central part of identity, and by looking at a world map which shows the distribution of both scripts and religions, the powerful correlation between religion and script becomes obvious. Examples include the correlations between, for instance, Orthodox Christianity and the Cyrillic script, Roman-derived Christianity and use of the Latin script, Islam and the Arabic script, Confucian religion and Chinese script, Brāhmī-Buddhist religion and the use of one of the Indian scripts, Judaism and Hebrew scripts, to name but a few.

However, religion is only one facet of identity conveyed by the use of a particular script, and there are numerous exceptions to this observation. For instance, Fula-speaking people in Africa are among the proponents of AJAMI or Arabic-based writing in sub-Saharan Africa, because they were among the first to be in contact with Islam. Ajami writing traditions are still dominant in many varieties of Fula, for instance in Guinea and Cameroon. Speakers of the Pular variety in Senegal, however, have broken with the tradition of Ajami writing and prefer a recently introduced Latin-based orthography, although they are still Muslims (Humery 2001; Humery-Dieng forthcoming). The reasons for this shift lie in the fact that in Senegal, Ajami writing was promoted by the Mourides, a Sufi brotherhood whose membership is mainly Wolof. In consequence, the Ajami tradition in Senegal became so strongly associated with its use for Wolof (called Wolofal) that speakers of Pular saw a Latin-based orthography as more appropriate for expressing their distinct identity.

Nevertheless, many endangered language communities come into contact with writing their own language for the first time through religious proselytizing, for instance by Christian missionaries aiming at bible translation and consequently engaging in literacy work. Therefore, the correlation between script and religion can still be very strong, even though the rise of the Latin alphabet through the global impact of information technologies and English sometimes makes it seem a 'neutral' script. Religious and identity aspects which influence the preference for one script over another may be very fine grained and not always deducible from general trends, and therefore, the careful investigation of identity-related issues is necessary prior to addressing the more technical sides of orthography development.

Scripts may serve to mark identity far beyond practical purposes. The Tifinagh script is an example of the powerful symbolism scripts or even single emblematic graphemes of them can carry. Tifinagh is an ancient Berber script whose actual use is probably negligible, despite the existence of a modern variety, Neo-Tifinagh. Yet, anybody remotely interested in Berber culture will have come across the grapheme *yaz*, prominently featured on the Berber flag (Figure 16.2).

Even if an old indigenous script is not used as the basis for a newly developed orthography, it is recommended to determine transliteration principles, if possible, and highly symbolic graphemes, may be graphically integrated, for instance by turning them into a logo. If at all feasible, the production of multigraphic documents should be considered, such as the bilingual and bigraphic Manding–English dictionary, which gives every LEMMA in the N'ko script used for the writing of a number of Manding varieties in Guinea, Côte d'Ivoire and Mali (Vydrine 1999); see Figure 16.3.[8]

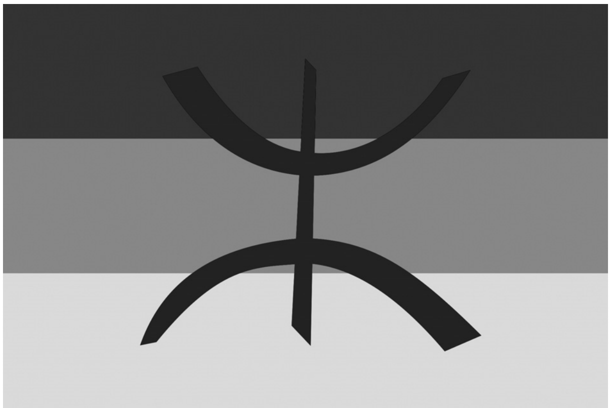


Figure 16.2. *Berber flag adopted in 1998 by the Amazigh World Congress, featuring the Tifinagh grapheme yaz http://commons.wikimedia.org/wiki/File:Berber_flag.svg (29 January, 2009).*

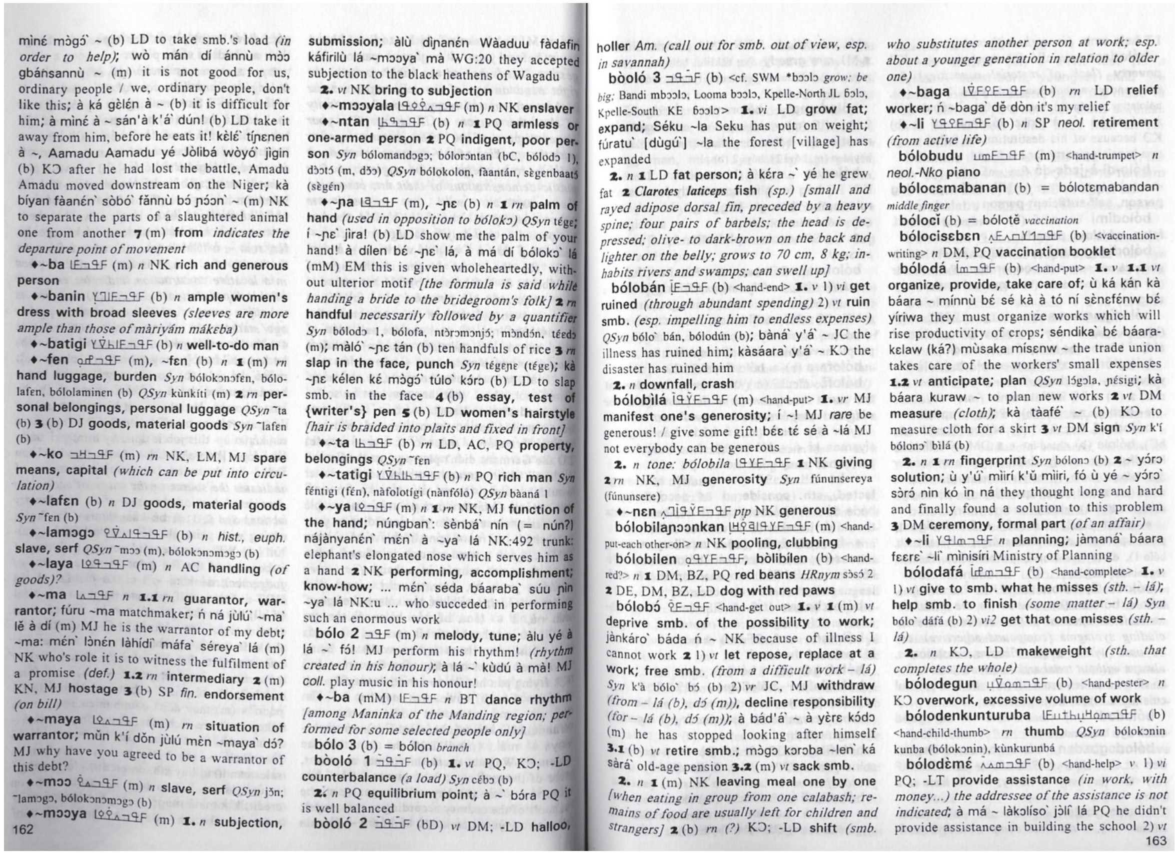
16.3.2 Are some scripts better than others for particular languages?

Another widely held belief about writing proposes that there is a natural proclivity of certain languages to be written with certain scripts. It is common, for instance, even among linguists, to voice objections against the use of an Arabic-based script for the writing of languages other than Arabic on the grounds that it would be impossible to represent the vowels of that language. Arabic has only three short vowels, and the role of vowels in written Arabic is smaller than that of consonants, reflecting the importance of CONSONANTIC roots and NON-CONCATENATIVE morphology in this and other Semitic languages. However, since the inception of writing, existing scripts have been adapted to suit the structures of very different languages repeatedly, often changing the type of writing system in the process.

To illustrate how the Arabic script may be used for languages with very different phonological and morphological properties, I present some examples. Most languages written in the Arabic script have more than three vowels, and many have consonants not found in Arabic. Three solutions to the problem of missing graphemes are available:

1. creation of new graphemes;
2. neutralization of contrasts of the spoken language in writing; or
3. appropriation of existing graphemes.

Hausa, an Afro-Asiatic language with a long Ajami writing tradition (Philips 2000), has adopted all three solutions. In contrast to Arabic, Hausa has five vowels. Vowel length is distinctive, as in Arabic. The short vowels /a/ and /i/ are written with the same diacritics as in Arabic,



mìnɛ́ mɔ̀gɔ́` ~ (b) LD to take smb.'s load *(in order to help)*; wò mán dí ánnù mɔ̀ɔ gbáńsannù ~ (m) it is not good for us, ordinary people / we, ordinary people, don't like this; à ká gɛ̀lɛ́n à ~ (b) it is difficult for him; à mìnɛ́ à ~ sán'à k'á` dún! (b) LD take it away from him, before he eats it! kɛ̀lɛ́` tíɲɛnen à ~, Aamadu Aamadu yé Jòlibá wòyɔ́` jìgin (b) KƆ after he had lost the battle, Amadu Amadu moved downstream on the Niger; kà bíyan fàanɛ́n` sòbó` fǎnnù bɔ́ ɲɔ́ɔn` ~ (m) NK to separate the parts of a slaughtered animal one from another **7** (m) **from** *indicates the departure point of movement*

♦**~ba** (m) *n* NK **rich and generous person**

♦**~banin** (b) *n* **ample women's dress with broad sleeves** *(sleeves are more ample than those of màriyám mákɛba)*

♦**~batigi** (b) *n* **well-to-do man**

♦**~fen** (m), **~fɛn** (b) *n* **1** (m) *rn* **hand luggage, burden** *Syn* bólokɔnɔfɛn, bólolafɛn, bólolaminɛn (b) *QSyn* kùnkíti (m) **2** *rn* **personal belongings, personal luggage** *QSyn* ˜ta (b) **3** (b) DJ **goods, material goods** *Syn* ˜lafɛn (b)

♦**~ko** (m) *rn* NK, LM, MJ **spare means, capital** *(which can be put into circulation)*

♦**~lafɛn** (b) *n* DJ **goods, material goods** *Syn* ˜fɛn (b)

♦**~lamɔgɔ** (b) *n* *hist., euph.* **slave, serf** *QSyn* ˜mɔɔ (m), bólokɔnɔmɔgɔ (b)

♦**~laya** (m) *n* AC **handling** *(of goods)?*

♦**~ma** *n* **1.1** *rn* **guarantor, warrantor;** fúru ~ma matchmaker; ń ná jùlú` ~ma` lě à dí (m) MJ he is the warrantor of my debt; ~ma: mɛ́n` lɔ̀nɛ́n làhídi` máfa` séreya` lá (m) NK who's role it is to witness the fulfilment of a promise *(def.)* **1.2** *rn* **intermediary** **2** (m) KN, MJ **hostage** **3** (b) SP *fin.* **endorsement** *(on bill)*

♦**~maya** (m) *rn* **situation of warrantor;** mǔn k'í dǒn jùlú mɛ̀n ~maya` dɔ́? MJ why have you agreed to be a warrantor of this debt?

♦**~mɔɔ** (m) *n* **slave, serf** *QSyn* jɔ̌n; ˜lamɔgɔ, bólokɔnɔmɔgɔ (b)

♦**~mɔɔya** (m) **1.** *n* **subjection, submission;** àlù dìɲanɛ́n Wàaduu fàdafin káfirilù lá ~mɔɔya` mà WG:20 they accepted subjection to the black heathens of Wagadu

2. *vt* NK **bring to subjection**

♦**~mɔɔyala** (m) *n* NK **enslaver**

♦**~ntan** (b) *n* **1** PQ **armless or one-armed person** **2** PQ **indigent, poor person** *Syn* bólomandɔgɔ; bólorɔntan (bC, bólodɔ 1), dɔ̀ɔtɔ́ (m, dɔ̌ɔ) *QSyn* bólokolon, fàantán, sɛ̀gɛnbaatɔ́ (sɛ̀gɛ́n)

♦**~ɲa** (m), **~ɲɛ** (b) *n* **1** *rn* **palm of hand** *(used in opposition to bólokɔ) QSyn* tɛ́gɛ; í ~ɲɛ` jìra! (b) LD show me the palm of your hand! à dílen bɛ́ ~ɲɛ` lá, à má dí bólokɔ` lá (mM) EM this is given wholeheartedly, without ulterior motif *[the formula is said while handing a bride to the bridegroom's folk]* **2** *rn* **handful** *necessarily followed by a quantifier Syn* bólodɔ 1, bólofa, ntɔ̀rɔmɔnjɔ́; mɔ̀ndɔ́n, tɛ́ɛdɔ (m); màló` ~ɲɛ tán (b) ten handfuls of rice **3** *rn* **slap in the face, punch** *Syn* tɛ́gɛɲɛ (tɛ́gɛ); kà ~ɲɛ kélen kɛ́ mɔ̀gɔ́` túlo` kɔ́rɔ (b) LD to slap smb. in the face **4** (b) **essay, test of {writer's} pen** **5** (b) LD **women's hairstyle** *[hair is braided into plaits and fixed in front]*

♦**~ta** (b) *rn* LD, AC, PQ **property, belongings** *QSyn* ˜fɛn

♦**~tatigi** (b) *n* PQ **rich man** *Syn* fɛ́ntigi (fɛ́n), nàfolotígi (nànfólo) *QSyn* bàaná 1

♦**~ya** (m) *n* **1** *rn* NK, MJ **function of the hand;** núngban`: sènbá` nín (= nún?) nájànyanɛ́n` mɛ́n` à ~ya` lá NK:492 trunk: elephant's elongated nose which serves him as a hand **2** NK **performing, accomplishment; know-how;** ... mɛ́n` séda báaraba` súu ɲìn ~ya` lá NK:u ... who succeded in performing such an enormous work

bólo 2 (m) *n* **melody, tune;** àlu yé à lá ~` fɔ́! MJ perform his rhythm! *(rhythm created in his honour)*; à lá ~` kùdú à mà! MJ *coll.* play music in his honour!

♦**~ba** (mM) *n* BT **dance rhythm** *[among Maninka of the Manding region; performed for some selected people only]*

bólo 3 (b) = **bólon** *branch*

bòoló 1 (b) **1.** *vt* PQ, KƆ; -LD **counterbalance** *(a load) Syn* cébɔ (b)

2. *n* PQ **equilibrium point;** à ~` bɔ́ra PQ it is well balanced

bòoló 2 (bD) *vt* DM; -LD **halloo, holler** *Am. (call out for smb. out of view, esp. in savannah)*

162

bòoló 3 (b) <cf. SWM *bɔɔlɔ *grow; be big;* Bandi mbɔɔlɔ, Looma bɔɔlɔ, Kpelle-North JL ɓɔlɔ, Kpelle-South KE ɓɔɔlɔ> **1.** *vi* LD **grow fat; expand;** Séku ~la Seku has put on weight; fúratu` [dùgú`] ~la the forest [village] has expanded

2. *n* **1** LD **fat person;** à kɛ́ra ~` yé he grew fat **2** ***Clarotes laticeps*** **fish** *(sp.) [small and rayed adipose dorsal fin, preceded by a heavy spine; four pairs of barbels; the head is depressed; olive- to dark-brown on the back and lighter on the belly; grows to 70 cm, 8 kg; inhabits rivers and swamps; can swell up]*

bólobán (b) <hand-end> **1.** *v* 1) *vi* **get ruined** *(through abundant spending)* 2) *vt* **ruin smb.** *(esp. impelling him to endless expenses) QSyn* bólo` bán, bólodún (b); bàná` y'á` ~ JC the illness has ruined him; kàsáara` y'á` ~ KƆ the disaster has ruined him

2. *n* **downfall, crash**

bólobìlá (m) <hand-put> **1.** *vr* MJ **manifest one's generosity;** í ~! MJ *rare* be generous! / give some gift! bɛ́ɛ tɛ́ sé à ~lá MJ not everybody can be generous

2. *n* *tone: bólobila* **1** NK **giving** **2** *rn* NK, MJ **generosity** *Syn* fúnunsereya (fúnunsere)

♦**~nɛn** *ptp* NK **generous**

bólobilaɲɔɔnkan (m) <hand-put-each other-on> *n* NK **pooling, clubbing**

bólobilen , **bòlibílen** (b) <hand-red?> *n* **1** DM, BZ, PQ **red beans** *HRnym* sɔ̀sɔ́ 2 **2** DE, DM, BZ, LD **dog with red paws**

bólobɔ́ <hand-get out> **1.** *v* **1** (m) *vt* **deprive smb. of the possibility to work;** jànkáro` báda ń ~ NK because of illness I cannot work **2** 1) *vt* **let repose, replace at a work; free smb.** *(from a difficult work – lá) Syn* k'à bólo` bɔ́ (b) 2) *vr* JC, MJ **withdraw** *(from – lá (b), dɔ́ (m))*, **decline responsibility** *(for – lá (b), dɔ́ (m))*; à bád'á` ~ à yɛ̀rɛ kɔ́dɔ (m) he has stopped looking after himself **3.1** (b) *vt* **retire smb.;** mɔ̀gɔ kɔrɔba ~len` ká sàrá` old-age pension **3.2** (m) *vt* **sack smb.**

2. *n* **1** (m) NK **leaving meal one by one** *[when eating in group from one calabash; remains of food are usually left for children and strangers]* **2** (b) *rn (?)* KƆ; -LD **shift** *(smb. who substitutes another person at work; esp. about a younger generation in relation to older one)*

♦**~baga** (b) *rn* LD **relief worker;** ń ~baga` dě dòn it's my relief

♦**~li** (b) *n* SP *neol.* **retirement** *(from active life)*

bólobudu (m) <hand-trumpet> *n neol.-Nko* **piano**

bólocɛmabanan (b) = bólotɛmabandan *middle finger*

bólocǐ (b) = bólotě *vaccination*

bólocisɛbɛn (b) <vaccination-writing> *n* DM, PQ **vaccination booklet**

bólodá (b) <hand-put> **1.** *v* **1.1** *vt* **organize, provide, take care of;** ù ká kán kà báara ~ mínnù bɛ́ sé kà à tó ní sɛ̀nɛfɛ́nw bɛ́ yíriwa they must organize works which will rise productivity of crops; séndika` bɛ́ báarakɛlaw (ká?) mùsaka mísɛnw ~ the trade union takes care of the workers' small expenses **1.2** *vt* **anticipate; plan** *QSyn* lɔ́gɔla, ɲɛ́sigi; kà báara kuraw ~ to plan new works **2** *vt* DM **measure** *(cloth)*; kà tàafé` ~ (b) KƆ to measure cloth for a skirt **3** *vt* DM **sign** *Syn* k'í bólonɔ` bìlá (b)

2. *n* **1** *rn* **fingerprint** *Syn* bólonɔ (b) **2** ~ yɔ́rɔ **solution;** ù y'ú` mìirí k'ǔ mìiri, fó ù yé ~ yɔ́rɔ` sɔ̀rɔ́ nìn kó ìn ná they thought long and hard and finally found a solution to this problem **3** DM **ceremony, formal part** *(of an affair)*

♦**~li** *n* **planning;** jàmaná` báara fɛɛrɛ` ~lì` mìnisíri Ministry of Planning

bólodafá (b) <hand-complete> **1.** *v* 1) *vt* **give to smb. what he misses** *(sth. – lá)*; **help smb. to finish** *(some matter – lá) Syn* bólo` dáfá (b) 2) *vi2* **get that one misses** *(sth. – lá)*

2. *n* KƆ, LD **makeweight** *(sth. that completes the whole)*

bólodegun (b) <hand-pester> *n* KƆ **overwork, excessive volume of work**

bólodenkunturuba (b) <hand-child-thumb> *rn* **thumb** *QSyn* bólokɔnin kunba (bólokɔnin), kùnkurunbá

bólodɛ̀mɛ́ (b) <hand-help> *v* 1) *vi* PQ; -LT **provide assistance** *(in work, with money...) the addressee of the assistance is not indicated*; à má ~ làkɔlíso` jɔ̀lí` lá PQ he didn't provide assistance in building the school 2) *vt*

163

Figure 16.3. *Sample pages from the Manding-English dictionary: bigraphic in Latin and Nk'o script (Vydrine 1999)*

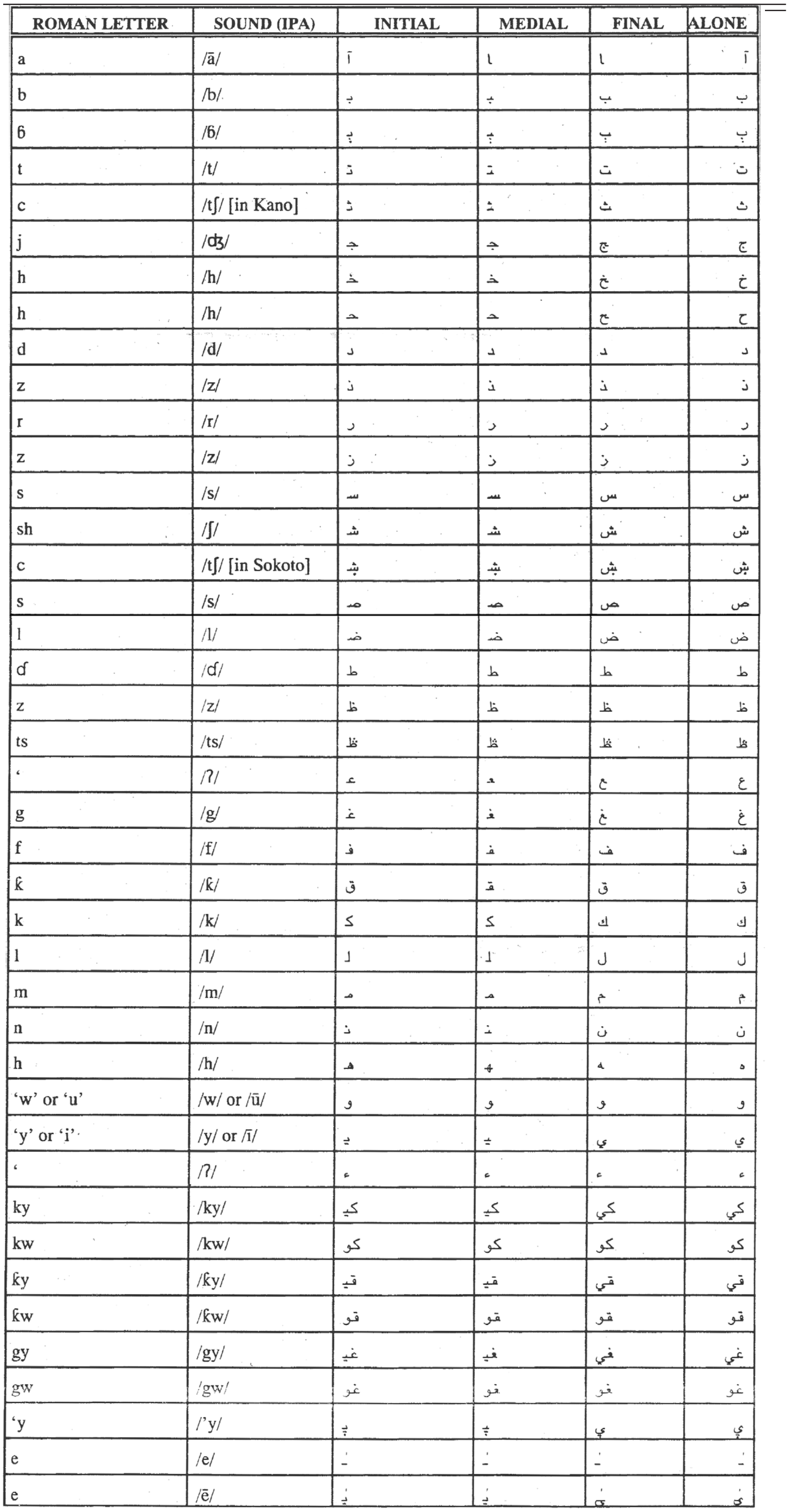
the *fatha* < َ > and *kasra* < ِ > respectively. Their long counterparts use the *'alif* < ا > and *yā'* < ي >. The phoneme /e/, not in the grapheme inventory of Arabic, is represented by a diacritic used in the Warsh tradition of writing the Qu'ran widespread in North and West Africa, a dot below or a vertical stroke above the letter. Its long counterpart is shown by an additional diacritic resembling a grave accent above the letter. Just like in Hausa, the Warsh grapheme indicates a phonetic [e] (see Table 16.2 for a chart of Hausa Ajami letters and their Romanized equivalents). The contrast between Hausa /u/ and /o/ and their long counterparts (signalled by a macron above the letter in Romanized Hausa) is neutralized, as both are represented by a symbol resembling the Arabic grapheme *damma* < ُ >.

The consonant inventory of many languages with Ajami writing is different from that of Arabic, and again, the same three different strategies can be observed. If the contrast is not neutralized, either a new symbol is created, such as < گ >, based on the letter *kāf* with an additional diacritic used to write /g/ in Persian. In Hausa, the *ghain* symbol < غ > is used to represent the same phoneme. Lameen Souag (p.c.) reports that the Arabic letter < ث > serves to write the affricates /dz/ as well as /ts/ in the endangered Songhay language Korandje of southwestern Algeria, an example of APPROPRIATION. Hausa, in contrast, represents /ts/ with a *ṭā* with three dots above (see Figure 16.4). A fourth adaptation strategy, also reported by Lameen Souag, is the use of a special diacritic that only specifies that a letter is to be pronounced like a similar, non-Arabic sound in the language being written, while leaving the exact pronunciation of that sound unspecified.

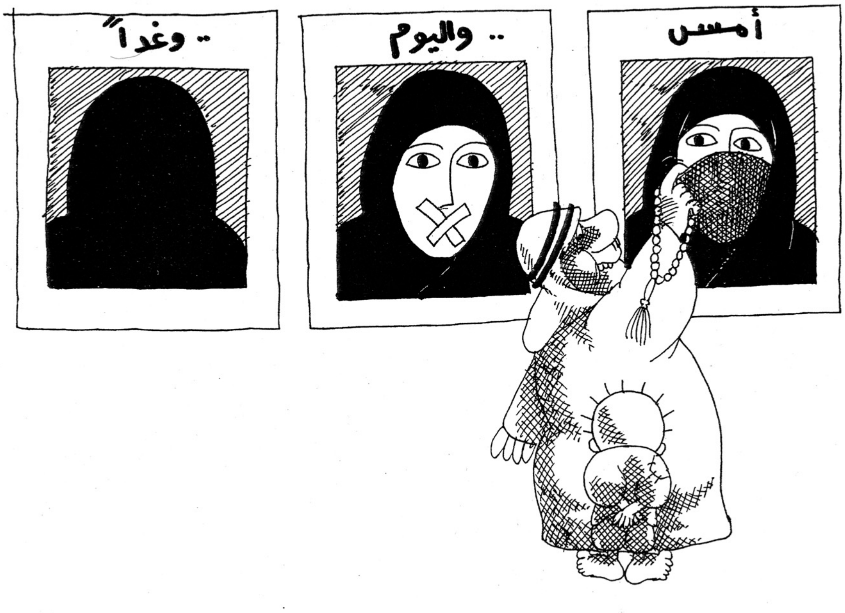
16.3.3 Practical matters

It is important to anticipate practical problems that might arise through the use of the chosen script. An important consideration concerns whether the script will be mainly read (often the case of endangered language literacies, as Trudell (2006) reports for three Cameroonian languages) or also actively written, and in which circumstances. The larger and less standardized the grapheme inventory, the less suitable is it for the use of new technologies in writing. Many manuscript cultures, for instance character-based writing systems such as the Chinese one, have a large inventory of characters that require complex input methods on a computer keyboard. The signs of these scripts can either represent a morpheme in Chinese or the language borrowing the script, or, through the phonetic form of the morpheme, a specific sound value (for instance the initial sound), and these differences will have a dramatic impact on the number of graphemes. Other scripts from manuscript cultures will typically contain graphemes that are not part of the Unicode standard (that determines the universal assignment of codes

Table 16.2. *Hausa Ajami chart after Philips (2000: 21f.)*



ROMAN LETTER	SOUND (IPA)	INITIAL	MEDIAL	FINAL	ALONE
a	/ā/	آ	ـا	ـا	آ
b	/b/	بـ	ـبـ	ـب	ب
ɓ	/ɓ/	ٻـ	ـٻـ	ـٻ	ٻ
t	/t/	تـ	ـتـ	ـت	ت
c	/tʃ/ [in Kano]	ثـ	ـثـ	ـث	ث
j	/ʤ/	جـ	ـجـ	ـج	ج
h	/h/	خـ	ـخـ	ـخ	خ
h	/h/	حـ	ـحـ	ـح	ح
d	/d/	د	ـد	ـد	د
z	/z/	ذ	ـذ	ـذ	ذ
r	/r/	ر	ـر	ـر	ر
z	/z/	ز	ـز	ـز	ز
s	/s/	سـ	ـسـ	ـس	س
sh	/ʃ/	شـ	ـشـ	ـش	ش
c	/tʃ/ [in Sokoto]	شٜـ	ـشٜـ	ـشٜ	شٜ
s	/s/	صـ	ـصـ	ـص	ص
l	/l/	ضـ	ـضـ	ـض	ض
ɗ	/ɗ/	طـ	ـطـ	ـط	ط
z	/z/	ظـ	ـظـ	ـظ	ظ
ts	/ts/	ڟـ	ـڟـ	ـڟ	ڟ
‘	/ʔ/	عـ	ـعـ	ـع	ع
g	/g/	غـ	ـغـ	ـغ	غ
f	/f/	فـ	ـفـ	ـف	ف
ƙ	/ƙ/	ق	ـقـ	ـق	ق
k	/k/	كـ	ـكـ	ـك	ك
l	/l/	لـ	ـلـ	ـل	ل
m	/m/	مـ	ـمـ	ـم	م
n	/n/	نـ	ـنـ	ـن	ن
h	/h/	هـ	ـهـ	ـه	ه
‘w’ or ‘u’	/w/ or /ū/	و	ـو	ـو	و
‘y’ or ‘i’	/y/ or /ī/	يـ	ـيـ	ـي	ي
‘	/ʔ/	ء	ء	ء	ء
ky	/ky/	كيـ	ـكيـ	ـكي	كي
kw	/kw/	كو	ـكو	ـكو	كو
ƙy	/ƙy/	قيـ	ـقيـ	ـقي	قي
ƙw	/ƙw/	قو	ـقو	ـقو	قو
gy	/gy/	غيـ	ـغيـ	ـغي	غي
gw	/gw/	غو	ـغو	ـغو	غو
‘y	/’y/	ۑـ	ـۑـ	ـۑ	ۑ
e	/e/	ـِ۬	ـِ۬	ـِ۬	ـِ۬
e	/ē/	يـ۬	ـيـ۬	ـى۬	ى۬



(Right to left) Yesterday, today, tomorrow: women's rights are stifled by conservative Arab elites (January 1985)

Figure 16.4. *Lebanese cartoon, and Chinese cartoon*
from www.al-akhbar.com/files/images/p24_20071210_pic1.full.jpg
from www.coe.tamu.edu/~kmurphy/writings/ptc90pap.html

to characters for their use with computers). This is the case for many Ajami scripts

It is not recommended to use characters on a keyboard that are not encoded in a Unicode standard, a computer, so there are two solutions if a transition from a manuscript culture of writing to word-processing (or text messaging on mobile phones) is desired: either the script is adapted so it uses only characters approved by the Unicode consortium,[9] or a proposal for a new script or character is submitted for approval. The latter is a time-consuming process only likely to be successful if the character or script is not an idiosyncrasy of one minority language, and it is therefore not a promising route for endangered languages. However, depending on the envisaged scope and function of literacy in the language, it may not be necessary to use computers to write it. Handwritten texts can be copied, scanned and disseminated, and local particularities of the manuscript culture can thus be preserved. This strategy may also be useful in contexts where computers are not widely available, or not equipped to handle complex scripts which are not contained in the regional Unicode subset of the area or which do not have the necessary fonts installed to display them. It is, however, generally advised to adhere to Unicode standards in order to cater for possible future developments that would make computers more accessible to the language community (see also Holton, Chapter 19).

In addition, a new and unexpected use of vernacular literacy sidesteps standardized and more formal literacies: text messaging. In many African situations, for instance, as observed by Stuart McGill (p.c.) and myself, text messages are the most common, if not the only, context in which local languages are used in writing. If it turns out that this register is going to be one of the predominant contexts for writing in an endangered language, this has a drastic impact on the inventory of graphemes available to be used for the orthography.

If the immediate use of the endangered language on computers is desired, some further considerations are in order. Most logographic scripts have complex interfaces for character input which will not be further discussed here. For alphabetic scripts, it is worth considering which keyboard(s) is/are standard in the areas in which the script is to be used. Although many linguists use keyboard mapping software to create tailored keyboards for specific languages or master other input methods, it should not be forgotten that in many areas of the world computers are only accessible in internet cafes and chat rooms where only standard keyboards will be available, and where users are not necessarily familiar with short cuts or the use of the character map etc. in order to insert characters into a document (e.g. it is not convenient to write diacritics using keyboards geared towards English). See Seifart (2006) for similar points.

Finally, the directionality of the selected script will dictate the flow of writing on the page. In addition, it will also influence conventions for picture reading: not just for the interpretation of sequences of pictures but also for expectations on their composition (for instance the location of an agent to the right versus the left of a picture: Dobel *et al.* 2007, Maass and Russo 2003). This factor should be taken into account when planning the creation or reuse of illustrations for publications in the endangered language, illustrated for Arabic and Chinese in Figure 16.4

16.4 Choosing an orthography

Philologists, linguists and educators have insisted for several centuries that the ideal orthography has a one-to-one correspondence between grapheme and phoneme. Others, however, have suggested deviations for such functions as distinguishing homophones, displaying popular alternative spellings, and retaining morpheme identity. If, indeed, the one-to-one ideal were accepted, the International Phonetic Alphabet should become the orthographic standard for all enlightened nations,

> yet the failure of even a single country to adopt it for practical writing suggests that other factors besides phonology are considered important for a writing system. (Venezky 2004: 139)

This section is concerned with the non-linguistic, linguistic and practical questions surrounding the development of an orthography once a script or writing system has been determined. The section begins by examining how orthographies, like scripts, may reflect a community's identity through the choice of particular graphemes, spellings, etc. These choices can either express proximity to an existing orthography by copying its conventions, or distance by using different graphemes and spelling norms from surrounding orthography traditions. Section 16.4.2, on design considerations, outlines some fundamental linguistic and psycholinguistic principles on the relationships between sounds and graphemes, shallow versus deep orthographies, etc. Section 16.4.3 discusses the practical impacts of particular choices of, e.g. graphemes, diacritics, digraphs, etc. on the production of written materials and scope of use of the orthography.

16.4.1 Orthography and identity

It is not only scripts which signal proximity to or distance from surrounding religions and ethnic and/or linguistic groups; orthographies, too, express similar aspects of identity. The retention of graphemes already in use in the speech community or in nearby literacies will situate the orthography within their tradition. This may be acceptable to the endangered language community, or it might be seen as intolerable. If choices are not constrained by higher order decisions such as national or regional conventions, it may be necessary for acceptance to take speakers' concerns regarding the choice of particular symbols seriously. For instance, members of the Miraña speech community in South America insisted on choosing graphemes that were visually different from those of the neighbouring Bora group, as Seifart (2006) reports. The motivation to express a distinct identity through different graphemes is sociopolitical; the Mirañas are outnumbered by the closely related Bora and strive to maintain their own ethnic identity. A contrasting driving force underlies the use of <ʉ> to write a high central vowel in the different alphabets for Cameroonian languages of the Bamileke group (Bird 2001). The different Bamileke varieties do not have a unified orthography, yet the barred ʉ has become a symbol of cultural unity. Other Grassfields languages not belonging to the group write the high central vowel as <ɨ>. A new would-be standard orthography for the entire group retains <ʉ> although it does not conform to orthographic conventions.

Identity-related motivations may sometimes conflict with linguists' attempts to create an orthography with transparent and predictable grapheme inventories that are consistent with conventions for neighbouring languages, although (or maybe because) the latter would facilitate transfer of literacy skills. Similar issues hold for the spellings of individual words. While not all orthographies are committed to reflecting the etymology of words (see Section 16.4.2 below), specific items may be of particular cultural significance, and communities may insist on spelling these words according to different principles than others, or on spelling them to reflect folk etymologies. It is recommended to evaluate identity-related issues with members of the language community, bearing in mind orthographic systems in regional and national use, in order to avoid decisions that might result in the rejection of the orthography by the community or some members of it (which might lead to digraphia or multigraphia; see Section 16.2.2 above).

16.4.2 Design considerations for orthographies

It is widely assumed by linguists that the basis of the ideal orthography is phonemic. If this was the case, the main difference between a phonemic transcription and an orthography would be the inventory of symbols used; IPA symbols in the former, a different and potentially open-ended set of graphemes in the latter case. Writing and reading are, however, cognitive tasks that are very different from speaking and hearing, and rely to a much lesser extent on phonological recoding than orthography developers often believe. This being said, there are indeed orthographies that are very close to the phonology of the language written: so-called SHALLOW or SURFACE ORTHOGRAPHIES (of which a famous example is the Finnish orthography), but their existence owes as much to the number of phonological processes in the language as to orthography design.

At the opposite end of the spectrum are DEEP ORTHOGRAPHIES, of which English is a notorious example. These do not have a close correspondence to the phonological structure of isolated words, so that irregular phoneme–grapheme relationships are common. (The properties characterizing connected speech are generally not encoded by orthographies.) However, even shallow orthographies can deviate from the often invoked ‘phonemic ideal’ on principled grounds, specifically when faced with capturing phonological processes at the word level. It may be desirable not to represent their pronunciation exactly but to preserve the identity of morphemes in written form. In English, for instance, the plural morpheme is written <s> despite voicing contrasts in, for example, /kæts/ <cats> vs. /dɒgz/ <dogs>, so that the identity of the plural morpheme /-s/ is preserved even in contexts of neutralization. In Dutch, the preference is to match the pronunciation difference in writing, hence <reizen>, ‘travel’ vs. <reijst>, ‘travels’. It may be useful to let speakers decide in these contexts: they may be more alert to the existence of certain phonological processes than others, for instance. Speakers of the endangered

Austronesian language Bierebo, spoken in Vanuatu, systematically failed to apply the phonemic principle underlying the newly created orthography in one specific case: the language has HOMORGANIC PRENASALIZED STOPS, but following the phonemic principle it was decided not to represent the prenasalization orthographically. Nonetheless, native speakers intuitively do represent it when speaking, particularly intervocalically, e.g. /kulbembe/ 'butterfly' is spelled <kulbebe> (Peter Budd, p.c.). However, if an orthography is being developed primarily in order to provide language-learning materials for non-native speakers, knowledge of such principles may need to be encoded through the orthography.

While it is commonly assumed that it is better in an orthography to overspecify than to underspecify, UNDERSPECIFICATION (or the conflation of several phonemes into one grapheme) can be a powerful tool for the creation of a PANDIALECTAL orthography in the case of unstandardized and internally diverse speech varieties. Seifart (2006) reports the case of the Austronesian language Sasak, spoken in Indonesia. The practical orthography of Sasak as proposed by Peter K. Austin only contains five vowels, although some dialects have up to eight vowel contrasts other dialects have fewer. A unified orthography is here seen as outweighing the fact that the under-differentiation of vowels in some dialects leads to the existence of homographs. Since semantic and collocational cues are available in reading, this does not render the orthography less effective.

While it is important to decide at a SEGMENTAL level whether an orthography should systematically give preference to morpheme identity vs. representation of some phonological processes, it is a matter of debate whether it should encode SUPRASEGMENTAL properties of speech such as distinctive stress or tone. Many orthographies notate tone using diacritics, numbers or other graphemes, depending on regional convention. Yet studies of some recent orthographies of complex tone languages question the effectiveness of tone notation, for both writing and reading. In a survey of tone-marking conventions for African languages, Bird (1999) reports the result of an evaluation of tone writing in the Cameroonian Grassfields language Dschang. Tone in this language is written with diacritics; the low tone is not marked. The tonal conventions result in 56 per cent of written vowels bearing a tonal diacritic. However, most speakers of Dschang do not master the tone notation conventions at all, which may be due to the large number of tone SANDHI and to the additional presence of grammatical tone in the language. Similar problems in writing tone have been reported from other languages. It may be crucial to carry out a detailed investigation of tone initially, in order to understand its functional load in the language. It can then be decided what its functional load in an orthography might be (not the same), and a pilot orthography can be tested with community members for both reading and writing.

The issue of the functional load of graphemes is of more general relevance, and pertains to the representation of low-frequency phonemes in the orthography. In Jalonke for instance, the VELAR NASAL /ŋ/ is only contrastive in medial position. It is the only nasal to occur morpheme-finally. My choice was to allocate this sound its own grapheme (see Figure 16.1), but an alternative would have been to represent it with the same grapheme as was used for the ALVEOLAR NASAL /n/, (where the POINT OF ARTICULATION is the result of REGRESSIVE ASSIMILATION and hence not phonemic anyway). It would then have been written <nga>, while <daŋŋεε>, where the velar nasal occurs at a morpheme boundary followed by the definite suffix, /-εε/, would be written <dannεε>. Instead of leaving this sometimes difficult choice to the linguist, the speech community can be involved in the decision-making process. McGill and Wade (2008) not only present clear guidelines for their proposed orthography of the endangered Benue-Congo language Cicipu of Nigeria, but also explain their choices so that the speech community can accept or reject parts of it on informed grounds. For instance, they present two possibilities for writing [tʃ]: either <c>, as in the contact language Hausa, or <ch>, as in the official language English.

Other design considerations concern the representation of GEMINATION and vowel length. Guidelines on the official Māori orthography, for example,[10] determine not only that long vowels are written with a MACRON diacritic above the vowel (<āhua>), but also that they are not written when their appearance at morpheme boundaries would result in an extra-long vowel. Morpheme boundaries thus remain GRAPHOTACTICALLY intact by representing two vowel graphemes instead of vowel plus macron. In the case of <ā>, when it combines with a base ending in <a>, <aa> is written: e.g. <whaka> + <āhua> becomes <whakaahua>, not *<whakaāhua> or *<whakāhua>. As the Māori example also illustrates, conventions for determining word boundaries (often said to be an artefact of writing in the first place) and how to write complex words need to be established and explained.

Finally, a SORT ORDER must be established to specify in which order the graphemes will follow each other, e.g. in a dictionary. All these tasks are not isolated technical problems but relate to the issues laid out above, since they all serve the potential purpose of signalling through borrowing or preserving, abolishing or innovating features from existing orthographies to position social practices in a multidimensional network.

16.4.3 Practical matters

The practical issues mentioned in Section 16.3.3 do not only hold for script choice, but also for orthographies. In addition, some more specific reflections are in order when developing a script. Directly related to its usability is the question of how its graphemes can be typed on a computer keyboard. Regional differences such as the British and American QWERTY versus the French AZERTY versus the Latin American QWERTY have an impact on how ergonomically glyphs can be typed, depending on where on the keyboard they are located (if they have keys allocated at all). The use of computers, not just to type glyphs but also to manage databases etc., as well as the involvement of Unicode, also require some attention to grapheme–glyph correspondences, the use of digraphs, punctuation marks, etc. It is crucial to select the correct character (i.e. not just a form resembling the intended grapheme but with the correct properties and semantics, e.g. a letter not a punctuation mark or a numeral) to represent the intended grapheme out of the huge inventory of 96,000 Unicode characters, and to find the correct upper and lower case matches for it.

For instance, in the past, it has been a regional convention in Côte d'Ivoire to use punctuation marks to signal tone (Bird 2001). Punctuation marks are also frequently used to encode vowel length, e.g. <:> in IPA, and in some orthographies the glottal stop [ʔ] becomes <?> for ease of typing (e.g. the use of <!> and <#> to encode click sounds in some Khoisan language orthographies). These practices do not conform to Unicode, and using them would mean that these marks are not considered part of a word by computers, but rather to signal a word boundary, as they do in major languages like English or French. This causes major problems for spelling checkers, concordancing software, internet search engines or sorting entries in a dictionary database, and so on. Unicode also has a fixed inventory of combining diacritics (i.e. diacritics that 'fuse' with the character they modify) and non-combining diacritics. It is not recommended to use characters that are not part of the Standard, such as characters in the Private Use Area.[11] In light of the growing importance of computers, mobile phones and other electronic devices, it is advisable to follow Unicode-dictated and other technical consideration even if the planned orthography is intended for a currently manuscript culture. While it is not necessary to write aided by a computer or a mobile phone, it is certainly short-sighted to exclude the future use of such devices through selecting non-standardized non-Unicode characters etc. (For more on the use of new technologies with endangered languages, see Holton, Chapter 19.)

16.5 Conclusion

> the script of a language, usually considered an interchangeable exterior form, works as a potential factoring its development, because, like writing systems and spelling conventions, it is perceived by the speech community as important. Since language is a mental *and* a social fact, this in itself causes writing to have an impact on language. (Coulmas 2003: 240; emphasis added)

This chapter has placed orthography development within a wider context of language ecology, where written language use is seen as one of the many registers available to communities and individuals, not just to convey and decode messages but to mark their social, religious, historical and/or linguistic identity. In the case of multilingual speech communities, these different features of identity can be associated with different spoken and written languages and/or writing systems or orthographies. It is therefore impossible to reduce the task of orthography development to a practical endeavour that requires clearly delimited linguistic or pedagogical expertise only. Rather, it has been argued that the parameters governing the selection of a writing system, script and orthography that constitutes the best fit for a community are multiple and multidimensional. Therefore, the different steps involved in the complex task of assessing the potential use of writing in and for an endangered language, and then devising an orthography, go beyond the capacity of one field linguist working single-handedly. Instead, they should be envisaged as a collaborative and multidisciplinary enterprise in close consultation with speakers of the endangered language throughout the process.

Graphization is impossible to achieve as an add-on to a linguistic documentation project unless sufficient time and resources are set aside for orthography development, to avoid short-lived and tokenistic outcomes. It can only be hoped that funding bodies which focus on the documentation of endangered languages and exclude measures such as orthography development from their scope will in the future become more sensitive to the pressing need to derive useful products from language documentation that can be of direct benefit for the communities themselves.

Notes

1 This chapter uses the term ‘contact language’ to mean any language of wider communication, a wider definition than that in Chapter 5.

2 Logographic scripts function very differently from syllabic and alphabetic ones, which create conventionalized (albeit very different) relationships between sounds and graphemes and rely on the category of word. Since logographic scripts for endangered languages are mainly confined to East Asia and are beyond the scope of my own research, they are not covered here. Syllabic scripts are similar to alphabets in that they have a varying degree of correspondence with the sounds of the language, but differ in the basic unit they assume, the syllable. Since my own experiences are with alphabetic orthographies I do not address syllabic orthographies in detail, although their design principles are close to those for alphabets.

3 As the different language names in two of the cited cases demonstrate, it is a delicate and controversial issue whether varieties with different

multigraphic, national, and/or religious affiliations are to be regarded as one language or two. This issue is not independent of the graphic traditions associated with them, since writing systems are markers of identity and languages are not purely linguistic entities but constructs relying on shared identity according to a number of social, political, historical, religious, etc. factors.

[4] Throughout the chapter I write graphemes between < >, phonemes between / /, and phones between [ ].

[5] Description of Baïnouk is currently under way, and only a preliminary phonological analysis is available. Specifically, the question of whether degree of aperture is distinctive for all vowel pairs, particularly [i] and [u] and their open counterparts, is still undecided. Therefore, only a phonetic notation is used here, although it does not represent a narrow phonetic transcription of the attested vowel values.

[6] This problem of corpus planning preceding prestige planning and status planning is addressed by Sallabank, Chapter 14.

[7] Skuttnab-Kangas and Phillipson (1995) concede that the notion of 'mother tongue' may be problematic and therefore suggest that an individual can have at least two mother tongues. This suggestion only reinforces the inadequacy of the term.

[8] An even more inclusive but also tremendously time-consuming solution for this dictionary would have been to also include Ajami representation of the lemmata, as Ajami writing is also attested for Manding (Vydrine 1998).

[9] see www.unicode.org/ (1 March, 2010).

[10] www.tetaurawhiri.govt.nz/english/pub_e/conventions2.shtml (26 January, 2009).

[11] A wealth of information and guidance is available at the homepage of the consortium (www.unicode.org).