

1 Creating Orthographies for Endangered Languages

Mari C. Jones and Damien Mooney

1 Introduction

The creation of new orthographical systems is often considered a key component of language revitalization efforts. The ability to encode an endangered language can facilitate the implementation of literacy programmes, which aim to reverse language shift by facilitating teaching and learning. Due to the fact that many endangered languages are unwritten, language planning involving orthography development (or graphization) is seen as ‘an essential prerequisite for many activities in favour of their maintenance and revitalization, such as dictionary writing, curriculum development and the design of language-teaching courses’ (Lüpke 2011:312). Additionally, in speech communities that are fragmented dialectally or geographically, a common writing system may enhance the status and prestige of an endangered language and may help create a sense of unified identity. Graphization can, in a sense, legitimize an endangered language owing to the widespread belief that a language must be written in order to be considered a language: ‘the existence of a written form almost lends mythical qualities to a language’ (Lüpke 2011:320; cf. ‘grapho-centrism’, Blommaert 2004).

Despite these apparent advantages, when a language is endangered, creating an orthography can bring with it a great number of challenges. The act of creating a new writing system requires the orthographer to consider a complex array of linguistic and, crucially, extra-linguistic factors. The trained linguist will at first be concerned with the type of script to adopt and with the level of linguistic structure that the orthography should represent, as well as with other issues such as establishing word boundaries and marking suprasegmental features. Secondly, they will also be required to take into account a wide variety of socio-political, psychological, and practical issues: the role of the native speaker in graphization, ideological distance from surrounding languages, knowledge of a dominant language’s orthographical conventions, cognitive and pedagogical issues involving reading, the existence of legacy orthographies, and technological reproduction. Many of these linguistic and extra-

linguistic factors differ in scope in situations involving endangered languages when compared with unwritten languages more generally (cf. Cahill and Rice 2014), and this volume will debate critical questions in this regard, with specific reference to obsolescent and revitalizing varieties.

In this introduction, we discuss the linguistics, sociolinguistics, and ideology of orthography development for endangered languages. We begin by discussing one of the central aims of orthography development, the creation of a literacy programme (Section 2), before broaching the issues involved in the standardization of endangered languages (Section 3), a process often considered to be a prerequisite for graphization. The discussion of linguistic (Section 5) and extra-linguistic (Section 6) factors involved in the development of new orthographies provides a detailed overview of the primary frameworks and issues involved. These issues are subsequently debated by the contributors to this volume (Section 7), focusing on a wide variety of case studies from around the world.

2 Literacy for Endangered Languages

Local literacy is often considered to be a ‘powerful tool’ for promoting and validating the use of endangered languages undergoing language shift (Grenoble and Whaley 2006:186), and therefore it can be seen to alleviate the pressure to speak and use the language of wider communication. Language revitalization efforts are commonly focused on the development of literacy programmes for endangered languages, as such programmes are seen to raise the status of the language involved, to modernize the language for use in a variety of social domains, and to facilitate school-based revitalization efforts (Grenoble and Whaley 2006:102). Literacy programmes for obsolescent languages are dependent on the existence of a developed orthography, and for that reason orthographies are essential for language revitalization models, which emphasize the effectiveness of literacy in reversing language shift. There are, of course, many other essential components of effective literacy programmes, in addition to orthography development: ‘other aspects include developing primers, teaching reading and writing, facilitating a variety of reading materials, training teachers, holding writer’s workshops, and a host of other activities’ (Rice and Cahill 2014:3). Therefore, the process of graphization does not in and of itself facilitate literacy or language revitalization – it must be integrated carefully into the larger process of corpus planning for endangered languages (Lüpke 2011:312–313).

Endangered languages are, by definition, subordinate and dominated by languages of wider communication, the majority of which have their own well-established orthographies. When literacy in the language of wider communication is highly valued, the development of literacy programs for the endangered

language can raise its status and prestige, encouraging and validating its use in a wider variety of domains. It is also possible, in situations where the speakers of an endangered language are not literate in the dominant language, that the acquisition of literacy can facilitate the acquisition of the dominant language, thus accelerating language shift, contrary to the central aims of language revitalization efforts (Grenoble and Whaley 2006:102). In the context of language revitalization, however, speakers of endangered languages are frequently at least partially literate in the language of wider communication. In this scenario, the development of a local literacy programme will bring the endangered language into direct competition with the dominant language and with its domains of use. This competition is absent for languages that are not threatened with extinction, and endangered language situations ‘differ fundamentally from those where there is no encroaching or dominant language of wider communication’ (Grenoble and Whaley 2006:126). Despite these possible drawbacks, Grenoble and Whaley (2006:102) argue for multiple literacies in language revitalization situations – that is, literacy in the endangered language and in the language of wider communication. The promotion of multiple literacies must, however, reduce competition between the dominant language and the endangered language by creating specific literacy contexts and domains of use for the latter.

The potential role of literacy in the endangered language speech community must be fully evaluated before orthography development and the creation of literacy programmes (Lüpke 2011:312), because existing domains of use and contexts for writing are often occupied by the language of wider communication. It is therefore necessary to identify an ‘ecological niche’ (Lüpke 2011:313) for writing or to create a context for written forms of endangered languages so that these can compete with the languages of wider communication: ‘this claim is supported by the work in local literacies which shows that literacy programs succeed when they are perceived as needed by the community and when the acquisition of literacy has some direct application to life in the community’ (Grenoble and Whaley 2006:127). Literacy materials in endangered languages (‘endographic traditions’) that replicate materials already available in the dominant language (‘exographic traditions’) are often viewed positively by the speech community in the first instance because they place both languages on equal footing, but Lüpke warns that such materials ‘have little or no practical use because established exographic traditions pre-empt the introduction of endographic ones for the same functions’ (2011:319). Only when a niche can be identified or a context can be created for endographic writing systems should orthography development and literacy programming begin. If such contexts cannot be identified or created, we must question the necessity of local literacy. Often, the obvious contexts for use of the endangered language in its written form are those pertaining to local traditions or local

culture, but Grenoble and Whaley (2006:129) advise that, for literacy efforts to be successful, they must extend beyond traditional cultural domains. They suggest encouraging, for example, writing for personal needs (e.g., writing diaries, writing notes, making lists), as this practice requires a personal investment and commitment on the part of the writer (and of any potential recipients) to use the local language in these contexts.

Several types of literacy programme have been proposed and should be considered before pursuing such a programme, and before an orthography can be developed. The concept of ‘traditional literacy’, which has as its central goal the teaching of reading and writing, is associated with literacy in languages of wider communication and with ‘formal, Western-style’ education (Grenoble and Whaley 2006:103); it is thus too restrictive and inapplicable to endangered language and revitalization contexts. We must consider understandings of literacy more relevant to the context of language revitalization (adapted from Grenoble and Whaley 2006:103–111):

- (i) *Autonomous literacy*: Autonomous literacy posits reading and writing as technical skills, independent of the social context and culture in which they are used (see Goody and Watt 1963). Within this model, a binary distinction is drawn between preliterate societies and literate societies, and literacy is viewed as a vehicle for social and cognitive change. Preliterate societies are considered underdeveloped (socially and cognitively), whereas literate societies are considered developed and modern: ‘on a cognitive level, it is argued that literacy is necessary for the cognitive development of certain skills, such as scientific reasoning, logic, abstract thinking, and the ability to distinguish between literal and metaphorical meanings’ (Grenoble and Whaley 2006:105). The view that, in order to become socially and cognitively equal, preliterate societies must become literate has since been discredited (see Goody 1987), but the situation remains that illiteracy is frequently stigmatized, underlining the potential importance of literacy programmes to status planning for endangered languages.
- (ii) *Vai literacy*: The concept of Vai literacy was developed by Scriber and Cole (1981) during five years of research with the Vai people in Liberia. They argued that it is formal education that influences cognitive development rather than literacy in absolute terms. Scribner and Cole’s (1981) study demonstrated empirically that literacy is not a purely technical skill, refuting the claims of autonomous literacy models and showing literacy to be a social construct, or a social practice, which can be revised and re-examined within its social context. For them, literacy is not simply about learning to read and write; it also involves the acquisition of knowledge about how to apply these skills in specific contexts and for specific purposes.

- (iii) *New Literacy Studies*: New Literacy Studies (Street 1984; Barton 2006) emphasize the social and cultural context of literacy, focusing on literacy as a ‘collection of things people do rather than a cognitive condition or skill’ (Sebba 2011:36). Literacy is viewed here as a social practice that will necessarily vary depending on the social context in which it finds itself. These studies require extensive ethnographic research into endangered language speech communities, to establish the individual needs of such communities before establishing a literacy programme; the development of an orthography and teaching materials is not seen as sufficient to promote and maintain literacy. New Literacy Studies highlight the potential impact of literacy on cultural practice, such as oral performance, and the researcher is reminded that the introduction of literacy will undoubtedly face the challenge of altering social practice (Grenoble and Whaley 2006:110).
- (iv) *Functional literacy*: The concept of functional literacy is commonly associated with adult education programmes and with UNESCO’s world-wide literacy campaign: ‘though the UNESCO plan asserts that each individual literacy program must develop its own operational definition of literacy, it is clear that, for UNESCO, literacy involves both reading and writing’ (Grenoble and Whaley 2006:111). UNESCO’s campaign emphasizes the links between literacy and socio-economic advantage and views literacy as providing access to specific opportunities for development and growth.

Community involvement and the consultation of native speakers are essential elements of any successful literacy programme, including the development of new orthographies (see Section 6). Modern literacy models insist on a conception of literacy that is firmly embedded in the social context of a given endangered language; literacy must respond explicitly to the needs of the specific speech community involved. The most efficient way of ascertaining these needs is to involve members of the speech community in the development of literacy programmes, or, ideally, ‘literacy will be the product of a grassroots kind of movement, coming from within the community itself and involving community participation in all phases of development’ (Grenoble and Whaley 2006:103). An idealized ‘bottom-up’ literacy campaign may prove difficult, particularly in preliterate societies, leading to a need for outside intervention by trained linguists. The speech community should, however, always be afforded ownership of the literacy programme if it is to succeed; community support is the key component of language revitalization, including orthography development and the creation of literacy. The trained linguist should avoid at all costs the implementation of a literacy programme (including associated orthographies and pedagogical materials) that has not received community approval. Grenoble and Whaley (2006:126) summarize the core

criteria for assessing community support before the introduction of local literacy: (i) the usefulness of a literacy programme must be recognized and approved by traditional community members (e.g., elders, politicians, religious leaders); (ii) local contexts for literacy must be identified and approved by community members; (iii) there must be continued widespread use of the endangered language as a spoken language; (iv) there must be support for the maintenance of local literacy by (local) educational systems. Criterion (iii) reminds the language planner that the development of orthographies and of literacy programmes must not be pursued independently of larger status and corpus planning goals as part of language revitalization efforts. Widespread spoken use of a language reinforces and supports the development and implementation of literacy campaigns for unwritten languages, but in the context of revitalization, endangered languages are by definition undergoing a rapid reduction in speaker numbers. Literacy programmes must therefore be implemented as part of a larger revitalization programme, because ‘creating literacy will not, in and of itself, revitalize a language’ (Grenoble and Whaley 2006:126).

3 Standardization for Endangered Languages

Linguistic standardization is often viewed as a prerequisite for orthography development, and indeed, standard ideologies are frequently centred on the highly codified written form of language (Bradley 2005:1; Seifart 2006:285; Lüpke 2011:313; Sebba 2012b:59). For Fishman, ‘in the modern world, standard dialects are written languages and they have definite written conventions, as far as writing system, orthography and grammatical structure are concerned’ (1991:346). Standardization is deemed necessary before orthography development, for a variety of reasons: writing may become idiosyncratic without linguistic standardization, rendering it inaccessible to a critical mass of speakers (Grenoble and Whaley 2006:130); a standard writing system is often seen as necessary for language promotion and for the development of formal teaching materials (Sallabank 2013:170); a standardized orthography encourages language use in a wider variety of domains, potentially raising the language’s status at the community level (Sallabank 2013:172; Grenoble and Whaley 2006:154); standardization and a standard orthography may help create a sense of a unified identity for the speech community. These advantages, particularly those relating to the status and community perception of the language involved in standardization, are said to be even more crucial within the context of revitalization (Fishman 1991:347). Successful corpus planning, including orthography development, constitutes a ‘powerful tool’ towards achieving the goal of reversing language shift if the standardized orthography can be read and understood by a large number of speakers (Grenoble and

Whaley 2006:130). Despite these clear advantages, the process of developing orthographies for endangered languages is often controversial and fraught with various linguistic and ideological issues, particularly when the language involved is spoken over a large geographical area and exhibits high levels of dialectal variation.

Dialectal variation is frequently viewed negatively by lay speakers, because it may give the impression that the speech community is fragmented both linguistically and in terms of its identity. While standardization is often considered a solution to this problem, the selection of linguistic forms to serve as the basis of a standard language can emerge as a contentious and divisive issue, particularly in the context of language endangerment and revitalization; the standard language may be seen to privilege speakers of one dialect over others (Grenoble and Whaley 2006:102). When endangered languages exhibit high levels of diatopic variation, or when multiple dialects are seen to be in competition, several options are available during the process of linguistic and orthographical standardization: (i) the unilectal approach, or selecting one particular dialect for standardization; (ii) the dialectal approach, or creating multiple standards; (iii) the multilectal approach, or creating a standard that contains linguistic features from a number of dialects; (iv) the common core approach, or creating a standard that emphasizes linguistic features common to all dialects.

The selection of one particular dialect to form the basis of the standard language and orthography is known as the unilectal, or ‘reference dialect’, approach to standardization: a single dialectal variety is promoted as the form of language to be used in formal and written domains, and speakers of other dialects are required to ‘[work] out the equivalences with their own dialects or [to learn] the standard as a new variety’ (Sebba 2012b:110). The selection process is influenced by both geographical and social factors: ‘(a) the relative location (a central location may have fostered a widespread regional comprehension due to bidialectalism); (b) the number of speakers; and/or (c) an elevated level of prestige’ (Karan 2014:115). It is commonly the most prestigious dialect that is selected for standardization in the unilectal approach, and problems may arise when there is no ‘generally agreed-upon prestige dialect’ (Cahill 2014:12). The unilectal approach should not, in theory, create problems for orthography development, because systematic differences – those exhibiting a one-to-one isomorphic correspondence – between the standard and non-standard dialects can accurately be represented by a single orthography (Venezky 1970:264). Fishman notes that it is necessary, particularly in the context of language revitalization, to avoid selecting a ‘highly divergent dialect’ for the role of standard (1991:343). From a linguistic perspective, problems may arise when, for example, a single phoneme in the standard dialect has two reflexes in another dialect, but problems associated with the unilectal

approach are normally social in nature. It is of utmost importance that the dialect selected for standardization be accepted by speakers of all dialects, if it is to succeed. If not, speakers of dialects that closely resemble the standard may be viewed as having an unfair advantage (Fishman 1991:343); resistance to the standard may occur if the written form is seen to depart too much from spoken varieties (Grenoble and Whaley 2006:130).

When linguistic differences between dialects impede mutual comprehension, or for ideological reasons, a ‘dialectal’ approach may be adopted: multiple standard dialects and corresponding orthographies are created during the standardization process. There are several drawbacks to this approach, however. The dialectal approach is not encouraged in revitalization contexts because it ‘emphasizes’ linguistic differences and socio-cultural distance between groups, resulting in linguistic and social fragmentation’ (Karan 2014:115). Indeed, Marcellesi (1983:216) notes explicitly that polynomia is easier to implement in situations of language maintenance than when attempting to reverse language shift. This is because revitalization efforts commonly attempt to elevate speaker numbers by encouraging second-language learning, and while native speakers may have no problems using and interpreting multiple dialectal orthographies, ‘second language learners need a model to aim at’ (Sallabank 2013:173). By contrast, when the aim is to promote literacy among fluent speakers, the development of a ‘teaching orthography’, which may be modified to reflect the linguistic forms of various dialects, may provide additional support from a pedagogical perspective (see the ‘differential’ approach, Karan 2014:116–117). In practice, the dialectal approach is uncommon; it is rare to find dialects that are so different from each other that no common orthography can be created, because ‘dialects, by nature, are characterized by systematic phonological differences’ (Venezky 1970:264).

It is necessary to distinguish between a dialectal approach, where individual standard orthographies are created for different dialects, and a multilectal approach, where the orthography created contains linguistic features from several different dialects. From a linguistic perspective, the development of a multilectal orthography involves representing ‘the phonologies of many dialects of a language [which] are compared and accounted for in designing the orthography’ (Simons 1977:325). Allerton (1982) recommends that all potential phonemic contrasts should be represented in the composite orthography: if two contrastive phonemes of one dialect, for example, correspond to only one phoneme in another dialect, the multilectal orthography will require speakers of the latter to ‘represent in writing a distinction which they do not make in speech’ (Sebba 2012b:110). Thus, a multilectal orthography does not represent any one spoken variety of the language (Cahill 2014:13; Grenoble and Whaley 2006:152), but the integration of linguistic features from a range of spoken varieties can serve to foster a common identity for the speech

community at large. The creation of a composite orthography is therefore commonly motivated by socio-political or sociolinguistic factors (Berry 1968:741; Karan 2014:117). Indeed, decisions regarding the linguistic features to include in the composite orthography will often depend heavily on the relative prestige of and socio-political relations among the contributory dialect groups (Seifart 2006:285). Simons (1994) presents a multilectal orthography as preferable to a unilectal one because the latter requires a large number of speakers to rote-learn various aspects of the unilectal system. Grenoble and Whaley (2006:153) argue, however, that this recommendation may not always be practical in situations of language revitalization – such as, for example, when dialects exhibit different levels of vitality and the decision to use the ‘healthiest’ variety as the basis of a unilectal standard may be more effective. Additionally, the development of a multilectal orthography may instigate socio-political tensions relating to the over- or under-representation of features from the contributing dialects: ‘the apparently “neutral” question of how best to accommodate different varieties within a single orthography leads directly to issues of power and authority’ (Sebba 2012b:112).

A final option is the development of an ‘artificial’ or common-core orthography which involves developing a writing system that emphasizes structural features that are common to all dialects. This frequently involves a certain degree of historical reconstruction, using prior forms of the language in which dialectal differences were less pronounced. The development of a common-core orthography, also known as a ‘unilectal approach based on an artificial dialect’ (Karan 2014:117), is more often motivated by socio-political rather than linguistic factors (Venezky 1970:264). Common-core orthographies have had limited success, and many linguists do not recommend them in situations of language revitalization (Karan 2014:117; Bradley 2005:4; Venezky 1970:264): the perception of the new standardized system as artificial and inauthentic may cause speakers to reject the newly developed orthography.

It is worth emphasizing at this point that status planning for newly developed orthographies and newly standardized dialects of all kinds must convey clearly to the speech community that the standard dialect is an additional variety and that it does not aim to replace spoken dialects: ‘all dialects should remain valid in speech within their own traditional speech networks and communities’ (Fishman 1991:342). The domains of use for the new standard must be clearly communicated to all members of the speech community, and spoken dialects of the language must be presented as equally valid. This is often difficult, particularly when the standard dialect and orthography are propagated by, and associated with, the education system, implicitly raising the prestige of the standard dialect. The success of standardization will be dependent on a common consensus that it is not an attempt to regulate speech or spoken dialects. In theory, creating separate domains for the written standard and for

spoken dialects is possible with careful language planning that encourages bidialectal competence. Indeed the celebration of local speech forms goes hand in hand with the postmodern *Zeitgeist*, which has led to an ‘increasingly positive attitude to local, non-standard varieties [...] particularly in areas where bi-dialectalism is defined clearly’ (Melchers 1987:187). Promoting bidialectalism can alleviate concerns during the standardization process about the loss of linguistic diversity.

Standardization can have both positive and negative outcomes for endangered languages. As indicated above, encouraging the development of standard dialects in an effort to reverse language shift can paradoxically result in the loss of local language varieties. Local varieties may adopt non-local or standardized forms via the process of de-dialectalization or dialect death due to the elevated prestige attached to standard forms (Sallabank 2013:172). When a unilectal or skewed multilectal approach is adopted, the attribution of prestigious social values to one dialect at the expense of others can lead to social stratification because speakers will have unequal access to the standard form of language (Grenoble and Whaley 2006:155). In turn, this social stratification can cause problems for revitalization efforts: the standard language may only ‘appeal to those whose life-style is modern/urban, a life-style in which written, formal and extra-local communication are all crucial aspects of the total repertoire’ (Fishman 1991:349), leading to social inequality. Grenoble and Whaley (2006:155–156) note, however, that situations of language shift emerge originally from social and sociolinguistic inequality between dominant and endangered speech communities and that, while standardization can create inequality within the endangered speech community, this may be the price to pay for minimizing inequality at a macro level between members of the endangered language speech community and speakers of the language of wider communication.

We have already noted that the existences of a prescriptive standard can raise the status of an endangered language due to the commonplace ideology that ‘real’ languages are written, standard languages. Jaffe insists that ‘orthography is one of the key symbols of language unity and status itself’ and that ‘it is not only important to “have” an orthography, but it is crucial for that orthography to have prescriptive power – to be standardized and authoritative, like the orthographies of dominant languages’ (2000:505–506). On the basis that authority and prescriptivism go hand in hand, Sebba (2011:44) argues that there is socio-political and socio-cultural agreement that spelling should be invariant – that is, that the introduction of writing entails the introduction of ideologies about correctness and about what is ‘right’ and what it ‘wrong’ in terms of language use (Karan 2014:109). Indeed, Grenoble and Whaley (2006:154) note that native speakers place less emphasis on correct language use before a language is given a written form.