Strategies for Representing Tone in African Writing Systems: A Critical Review

Steven Bird
University of Edinburgh
steven@cogsci.ed.ac.uk

May 1998

Abstract

Tone languages provide some interesting challenges for the designers of new orthographies. One approach is to omit tone marks, just as stress is not marked in English (zero marking). Another approach is to do phonemic tone analysis and then make heavy use of diacritic symbols to distinguish the 'tonemes' (exhaustive marking). While orthographies based on either system have been successful, this may be thanks to our ability to manage inadequate orthographies rather than to any intrinsic advantage which is afforded by one or the other approach. In many cases, practical experience with both kinds of orthography in sub-Saharan Africa has shown that people have not been able to attain the level of reading and writing fluency that we know to be possible for the orthographies of non-tonal languages. In some cases this can be attributed to a sociolinguistic setting which does not favour vernacular literacy. In other cases, the orthography itself might be to blame. If the orthography of a tone language is difficult to user or to learn, then a good part of the reason, I believe, is that the designer either has not paid enough attention to the *function* of tone in the language, or has not ensured that the information encoded in the orthography is *accessible* to the ordinary (non-linguist) user of the language. If the writing of tone is not going to continue to be a stumbling block to literacy efforts, then a fresh approach to tone orthography is required, one which assigns high priority to these two factors.

This article describes the problems with orthographies that use too few or too many tone marks, and critically evaluates a wide range of creative intermediate solutions. I review the contributions made by phonology and reading theory, and provide some broad methodological principles to guide someone who is seeking to represent tone in a writing system. The tone orthographies of several languages from sub-Saharan Africa are presented throughout the article, with particular emphasis on some tone languages of Cameroon.

1 Introduction

Until recently, tone was generally not marked in the languages of sub-Saharan Africa, even though the vast majority of these languages are tonal (Nida 1964:26). In such languages, the pitch on an individual syllable can be contrastive, distinguishing lexical items and grammatical categories such as verb tense. Designers of new orthographies were often ignorant about tone and imported many assumptions from the orthographic traditions of European languages (Baker et al. 1982:5).[1] These imported assumptions can be seen in the sound-symbol correspondence, so that *g* and not some other symbol denotes the voiced velar stop. They can also bee seen in the use of digraphs and trigraphs, so that *tsch* denotes the affricate *tʃ* in the original German-based orthography for the Dschang language (Djoumessi 1957); today this segment is written *c*. The very *sound distinctions* to be represented are heavily influenced too, as can be seen in the traditional resistance to mark tone:

> In discussions of the establishment of a practical orthography, for innumerable African languages and probably for languages elsewhere as well, resistance to any indication of tone has been exceedingly common, and has often approached the level of irrationality. (Welmers 1976:34)

Carrying over the phonemic contrasts of a colonial language into indigenous languages has led to the over-representation and under-representation of the necessary contrasts, and this in turn has made writing systems difficult to learn (Sapir 1933). Ample anecdotal evidence suggests that the omission of tonal distinctions from the orthography of many tone languages is a barrier to fluent reading.[2] However, the experimental literature on reading of tone languages is ambivalent; see Bird (1998) for a survey.

With the advent of phonemic theory there came a widespread acceptance that orthographies should be phonemic, where there is a one-to-one correspondence between phonemes and graphemes. Accordingly, tone also had to be marked (Pike 1948:36f). Williamson states that 'if the language is a tone language then we need some method for indicating the tones so as to distinguish different words' (Williamson 1984:7f). Mfonyam echoes the same idea: 'We ruled out a priori any system that did not mark tone because we were dealing with tone languages' (Mfonyam 1990:23). The assertion that tone should be marked in tone languages does not have to be accepted uncritically, however. For some tone languages there may be abundant non-tonal clues to the identity of a word or of a grammatical function.[3] While Pike (1948:37) and Williamson (1984:44f) countenance this possibility, it is seldom explored in depth.

Together with the assumption that tone should be marked in tone languages are the concomitant assumptions that tone should be marked *phonemically* and *diacritically*. Phonemic tone marking involves identifying the inventory of tonemes – according to some variety of phonemic

[1] For a review of the history and current state of orthographies in sub-Saharan Africa, the reader is referred to (Tucker 1971, Hartell 1993). Comprehensive surveys of the languages themselves are also available (Welmers 1973, Bendor-Samuel 1989), and for an introduction to the tone systems, see (Odden 1995).

[2] Pike (1947:252), Longacre (1964:132), Nida (1964:27), Klem (1982), Awedoba (1990:34), Koffi (1994:52)

[3] This point can be trivially exemplified for English. Consider two words which are spelt the same but pronounced with different stress, such as *ˈprotest* (noun) versus *proˈtest* (verb). Although the pronunciation difference is not represented, this is unlikely to cause problems since nouns and verbs are normally found in different parts of a sentence. Grammatical context cues the correct interpretation.

theory – and assigning a grapheme to each. Note that the same grapheme might be used for more than one toneme, and the *absence* of any tonal grapheme in a given position might also represent a toneme (a ‘zero grapheme’). Depending on which variety of phonemic theory is used, a contour tone may be a distinct toneme from the level tones, even though the contour might be trivially analysable as a sequence of level tones. These tonemes then form the basis of the tone marking system. Morphophonemic alternations may be fully spelled out: ‘tonemes substituted in morphology or syntax or sandhi should be written as pronounced’ (Pike 1948:37). In what follows, this phonemic marking, along with schemes that come closer to phonetic transcription, will be referred to as shallow marking schemes, for reasons that will become clear in §7.

The second assumption, referred to above, is that tone should be marked diacritically (Williamson 1984:41, Pike 1948:38), whether by accent marks as is the long-established practice in Africa (International African Institute 1930:14); by punctuation marks before each word, as in many Ivorian languages (Hartell 1993:124ff); by the recycling of underutilised consonant graphemes to mark tone (see §4.1); or by numbers as in many languages of Latin America, especially Mexico (Pike 1948:95ff).[4]

The goal of this paper is to survey and critically evaluate a variety of potential and actual tone marking schemes for African languages. The focus will be on the orthography as it is used in texts written by and for speakers of the language in question. In §2 the two extremes of shallow and zero marking are described. Then in sections 3-6 I survey various methods that have been used to mark tone in the languages of sub-Saharan Africa. We shall see that there is a veritable cafeteria of choices awaiting the designer of a new tone orthography. Following this, there will be a discussion of the contributions that phonology and reading theory have to make (§7). Here we shall that solutions to tone orthography problems will depend critically on the nature of the language itself. Finally, some methodological principles are discussed in §8.

Elsewhere, I review the experimental literature on reading African tone languages and report on my own experimental work (Bird 1998), and discuss a variety of sociological factors which impinge on orthography design and revision (Bird & Hedinger 1997).

2 Two Extremes: Shallow Marking and Zero Marking

Dschang, a Grassfields Bantu language of Cameroon, has an orthography which makes heavy use of tone marks. The Dschang tone orthography is founded on the phonemic, diacritic approach, and was introduced in the mid-1980s by Maurice Tadadjeu. An example text is shown in (1). High tone is marked using an acute accent and mid tone is marked with a macron. Low tone is unmarked.

(1) **Example text from Dschang:** *The squirrel and the dog.*

Kaŋ pɔ́ mbhʉ̄ é lelá’ ŋ́gɔ̄ mésō, ḿbú ńzīŋɛ́ ta’ enɔ. Pɔ́ lelá’ ńnāŋ tɛ eshʉ̄’ amɔ̄’ álī’í, ḿbɛ́ á ápa, ńdɔk ŋgʉ̄ɔ́ á ŋ́kā’ ŋiŋ njʉʉ́ a apumā. Pɔ́ le gē é tō á ḿba, ŋ́gɔ́ɔ mbɔŋ. Pɔ́ le gʉ̄ɔ́ tɛ ŋkó éwú, kaŋ á le mɛ́ mbhʉ̄ ŋ́gɛ: «Esó, pá’ meŋ ŋkʉɔ́k ḿbiŋ nzéŋ ńzāŋné lā, meŋ ɛ̄ kɔ́’ á áthʉ̄, é kāp, o gɔ́ á ésī ńnɔŋ mbɛ́tɛ́ ńnéŋ á ńtē á ápa».

[4] Note that Gudschinsky once recommended the use of tone numbers in Africa (Bolli 1978:17).

Based on over 4000 words of text, the density of tone marks in the orthography is 56%; *i.e.* over half the tone bearing units[5] carry an explicit tone diacritic. Only a handful of people have ever learnt to write the tone marks accurately. One man with some years of experience teaching the orthography reported that he usually writes the letters of a whole sentence or paragraph before going back to mark the tones. Although people can easily determine the tone of words in isolation by comparing them to a short list of key words that has been memorised, writing phrasal tone is an entirely different matter. The sentence must be recited several times in order to get the correct intonation. And this is not an isolated experience for Dschang:

> [Native speakers of Kasem (Gur, Ghana)] have difficulty in identifying and distinguishing between tones, as any exercise involving language teachers will show. ... Even those with training who can distinguish systematically between one tone and another are still wary about writing it. Many users cannot easily be taught to read tone correctly, not to mention writing it. (Awedoba 1990:38)

A major complicating factor in the case of Dschang is the phenomenon of tone lowering known as *downstep*, whereby a tone is lowered with respect to the material that has come before. In Dschang and many other tone languages, downstep has a *terracing* effect on the intonation pattern of a phrase, since 'lowered tones have the effect of "redefining" the pitch levels at which subsequent tones are realised' (Clements 1979:538). In Dschang, downstep regularly arises inside words, whether in isolation or in phrasal context. Orthographically, a downstepped high tone is represented with a macron and is called 'mid tone'. Currently, people are taught in lesson three of the reading and writing book[6] that in a high-mid-high word like ńdākté *gather*, the final tone is pronounced at the same pitch level as the preceding tone, resulting in the tone melody: [¯—]. Yet it is extremely difficult for speakers to generalise this idea so that it can be applied across word boundaries.

The Dschang tone orthography has another problem. Consider the words ŋ́káŋ *fry* and ŋ́kāŋ *choose*, which differ only in their tone. When said in isolation and in the context of a sentence, these two words have distinct tones, enabling the hearer to differentiate the meaning. However, in some situations it is not possible to represent the distinction in the orthography. An example is given in Table 1 for the present progressive. Note that the exclamation mark symbolises downstep. The unusual 'double downstep' can be seen between two high tones in Table 1, where it corresponds to the same pitch drop as is found in a high-low sequence.

Observe in the left column of Table 1 that both forms are written identically. The second and third columns show that the tone melodies are different, while the last column shows that the meanings are also different. Observe that the doubly downstepped high tone of ‼sí is represented orthographically as low tone. Now the high tone verb prefix ŋ́ must also be written as low, since it is at the same pitch as si. Finally, the tone contrast on the verb root, usually represented

[5] A tone bearing unit, in this context, is defined to be any letter which can potentially be marked with a tone diacritic. In Dschang, the tone bearing units are the vowels and the word-initial nasals (m, n, ŋ). The notion of tone density was introduced by Bird (1998).

[6] This is a book of twenty lessons for people already literate in French. People are taught in lessons two and three to mark tone on words in isolation by reference to a set of memorised keywords. Marking of tone on phrases is taught by example, rather than by teaching an explicit procedure. Tonal awareness is taught using whole-body exercises, standing and crouching to mimic voice pitch.

Table 1 Loss of Tone Distinctions in the Dschang Orthography

Orthography	*Pitch Contour*	*Phonology*	*Gloss*
meŋ á si ŋkaŋ	[–¯—_]	mèŋ á ꜜꜜsí ŋ́ꜜkáŋ	*I am choosing ...*
meŋ á si ŋkaŋ	[–¯——]	mèŋ á ꜜꜜsí ŋ́káŋ	*I am frying ...*

orthographically as the contrast between high and mid tone marks, cannot be represented without a symbol for lowered-low. However, introducing an extra symbol would not solve the problem, but simply perpetuate the effects of the original problem brought about by the conflation of low tone with doubly-downstepped high tone. It is not clear how this problem could be solved within the spirit of the present system without introducing explicit downstep graphemes into the orthography, a move which has never been tried as far as I am aware. A reader facing this construction (in texts containing reported speech) is faced with ambiguity, which is what the tone marking system was designed to avoid. Therefore, alongside the above complaint that this orthography has too much tone marking, we can add the further complaint that it does not have enough tone marking. This paradoxical situation is further illustrated by a comment of a Dschang literacy worker. He would like to see a reduction in the amount of tone marking. However, he also *adds* new tone diacritics (other than acute and macron) to texts he will read in public in order to avoid pauses caused by reading ahead for contextual clues.

Although the above impressions are largely subjective in nature, recent experimental work on Dschang has shown that its tone marking system does not support fluent reading (Bird 1998). Mother tongue speakers with a variety of ages and educational backgrounds, and having different levels of exposure to the orthography were tested. They read texts which were marked and unmarked for tone, then added tone marks to the unmarked texts. Analysis showed that tone marking degrades reading fluency and does not help to resolve tonally ambiguous words. Experienced writers attained an accuracy score of only 83.5% in adding tone marks to a text, while inexperienced writers scored a mere 53%.

In conclusion, we can say that people have not been equipped with an *effective procedure* for determining the correct tone marks on a written text. Since a dictionary can never list all the possible phrases, there is no end in sight for disagreements about the correct spelling of tone on phrases. Note also that the reading and writing book gives no instruction on phrasal tone marking, probably because there is no established method for teaching such tone transcription outside of a university phonetics laboratory. The result is a 'pepperbox' spelling system (Firth 1948:124), with accents peppering the page, burdening readers and writers with an unnecessary cognitive and mechanical load.

This negative experience with the phonemic, diacritic method of marking tone is being repeated in many languages. Ironically, we may end up where we began with no tone marking, as has happened for the seven officially recognised indigenous languages of Zambia:

> Tone, like vowel length, was accepted in all the seven languages as an important linguistic marker for signalling differences in meaning between two otherwise identical words. However, despite its functional importance, all the seven committees decided not to symbolize tone in the orthography. The basic argument advanced against its symbolization was that *experience had shown that diacritic marks, where employed, generally hindered rather than facilitated fluent reading*. It was further

> argued that ordinarily, even without visual representation, tone would be adequately signalled by context. (Kashoki 1981:169, emphasis mine)

So, perhaps a return to zero tone marking would not be such a bad idea. After all, one could cite the case of standard written Hebrew, where the diacritics (or *points*) are omitted even though these represent most of the vowels and help distinguish some of the consonants. Yet in the context of a sentence this unpointed form is usually adequate for word identification (Katz & Frost 1992:149). While many African tone languages have been written without any indication of tone, the practice has not met with universal success, as the following quotes attest:

> Present orthographic conventions [in Kasem] place the burden of homographs on the reader. He has to look for clues throughout the text to enable him to choose between different potential meanings. (Awedoba 1990:34)

> [A toneless orthography] encourages bad reading habits by forcing the beginner to read ahead for contextual clues, and then turn back to guess the meaning of earlier words. (Pike 1947)

In this article, my intention is not to take sides in the debate between zero and shallow tone marking, as if these were the only two options to consider, and as if a single policy would have universal applicability. Rather, in designing or revising a tone orthography, a more balanced approach is needed:

> Advocating the omission of diacritics to mark tone, stress and length may seem like linguistic heresy to some persons. Actually it is not. We simply need to recognize that for the speaker of a language it is not necessary to mark everything which is meaningful. In fact, the marking of such contrasts often seems unnecessarily cumbersome and awkward. ... Despite the practical and scientific validity for omitting the writing of tone, stress, and length in some languages, it should be said that in too many languages such distinctions are not written when they should be. And as a result readers of such languages have to stumble and guess unnecessarily. (Nida 1964:27)

Nida's statement is an eloquent summation of the approach I advocate here. We should be cautious about shallow tone marking for it is 'unnecessarily cumbersome and awkward' in many cases. We should be equally wary of zero tone marking since it may cause readers to 'stumble and guess unnecessarily'.

Tadadjeu & Sadembouo (1979:18) have suggested that 'it may be expedient to experiment in the marking of only "necessary" tones.' Though frequently proposed, this idea has never been systematically studied, to my knowledge. In fact, there are several ways to interpret the idea of marking 'only necessary tones'. In the ensuing sections I survey and evaluate a range of options for limiting tone marking. While this is done in the name of all tone languages, the reader should bear in mind that tone languages differ in their use of tone. Orthographic principles that work well in one language might perform poorly in another language. While a bewildering range of choices seems to confront us, we come to the 'smorgasboard' with strict dietary requirements that greatly constrain our choices. The constraints have to do with the function of tone in the language and the 'depth' of the tonal morphophonology, to be discussed in §7. Here, I simply wish to lay out the options in a neutral fashion. The survey is organised in terms of increasing abstractness. The least abstract approach is to adopt certain spelling rules, and it is this we shall turn to first.

3 Spelling rules to minimise tone marking

Probably the most widespread practice has been to omit one of the tone symbols so that a certain tone is simply represented by the absence of any mark. For example, in the current Dschang orthography already illustrated in (1), low tone is represented by the absence of any tone mark. The symbol omitted could be for the most frequently occurring tone (Pike 1947:222), where frequency is determined by reference to a wordlist. Although statistics from a large corpus of text would be preferable,[7] no sufficiently large corpus of text may yet exist in the language for this to be a realistic option.

A second method employs a rather different way of choosing which tone to omit:

> The tone(s) to be written are those which change the least. For example, if a high tone is downstepped to become a mid tone after a floating low tone, it is better to write low tone because it remains the most stable. (Wiesemann et al. 1983:156, my translation)

This requires us to determine which members of our tonal inventory manifest the least contextual variation. If low tone undergoes fewer changes than high tone, then low should be marked while high is unmarked. This is how tone was marked in Dschang in the mid-1980s. The tone orthography of Tikar (Southern Bantoid, Cameroon) is also based on this principle. Marking only the stable tones represents an interesting compromise between the ideal of writing tone phonemically and the desire to avoid representing morphophonemic alternations which change the visual appearance of a word in context. However, a consequence of marking only the most stable tones is that grammatical information signalled solely by a tone change may be lost.

A third approach is to mark the tone of a syllable only when it is different from the tone of the immediately preceding syllable (Williamson 1984:43). Welmers method for tone marking in Efik (Benue-Congo, Nigeria) uses this method (Essien 1977:166), and he also proposed this approach for Igbo. The idea may have originated in the 19th century work of Christaller. In this system, a sequence of identical tones receives just one orthographic mark which is placed on the first syllable of the sequence. So the first high tone of a sequence is marked with an acute accent and subsequent high tones are unmarked. A subsequent acute accent can then only stand for a *downstepped* high tone.

This approach may appeal for languages with tone spreading across word boundaries, when a word-initial syllable automatically acquires the final tone of the previous word. The process is so automatic that speakers are usually not aware that it has even happened. With this third tone marking scheme, automatic spreading would be trivial since iterating the same tone does not require extra diacritics. A further merit of this scheme derives from the claim that pitch *change* is more perceptually salient than absolute pitch *height*. Thus readers will see an orthographic mark at just the same point as they perceive a tonal event. However, there is a drawback to consider. Consider the Dschang examples in Table 2, paying particular attention to the underlined letters.

[7] As an example of why wordlist and text frequency are different, note that although ð occurs word-initially in less than thirty words in English, it is the most common word-initial consonant in continuous text (Lyons 1971:95). By large corpus in this context, I mean at least 5,000 words of text from a representative set of genres.

Table 2 Marking Pitch Change in a Tone Orthography for Dschang

Syllables	*After high tone*	*After low tone*	*Gloss*
3	...á è̱wɔŋkhʉ́	...à e̱wɔŋkhʉ́	*children*
2	...á ŋ̱kāŋ	...à ŋ̱́kāŋ	*cry*
1	...á pɔ̱	...à pɔ̱́	*pronoun*

It is easy to see that shorter words will be more variable in their appearance under this scheme. This means that grammatical words, which are typically short in Dschang and most other languages, will be the most variable in appearance, which is undesirable for the development of a sight vocabulary (see §7.3).

A fourth approach is possible when the tone occurring on a particular syllable is predictive of the tone pattern on the whole word. In Karang, an Adamawa language of Cameroon, tone on non word-initial syllables is never contrastive. Just the first syllable of each word is marked while later syllables are systematically unmarked (Ubels & Ubels 1986:6). The same approach has been successful in some Ivorian languages, such as the Kwa language Attié (Kutsch Lojenga 1986:60). In these languages, the tone on the first syllable is marked using a punctuation symbol which appears before the word. To avoid breaking up words, only the first syllable is marked for tone (Kutsch Lojenga, pers. comm. 1996). In the Ijoid languages of Nigeria, sentences divide into 'tone groups' in which the overall tone melody depends primarily on the first word of the group (Jenewari 1989:113). Marking only the first word for tone might be sufficient for readers to infer the melody of the whole tone group (Larry Hyman, pers. comm. 1998).

A fifth approach is to collapse some tonal distinctions while still marking others. Functionally or phonetically similar tones are marked using the one grapheme. For example, Kom, a Grassfields Bantu language of Cameroon, has been analysed as having nine tones, but only two tone diacritics are used. The correspondence between the tones and the diacritics is shown in Table 3 (Chia and Kimbi, 1992: 10; Jones, 1996: 4). The tones are high (H), mid (M), low (L), extra-low (XL), and certain combinations of these.

Table 3 Tone to Grapheme Mapping for Kom

Tone:	H	M	L	XL	HM	HL	HXL	LH	MH
Grapheme:			`	`		^	^		

Another approach which is orthogonal to the others works in languages which lack a vowel length distinction. Contour tones can be written across two vowels. If low tone is also omitted by default, then we have *â* becoming *áa* and *ǎ* becoming *aá* (Wiesemann et al. 1983:156f), and the number of diacritics is reduced from three to one. This has been successful in Anyi (Koffi 1994:57), a Kwa language of Benin. However, writing tone contours in this way will not work if readers associate vowel doubling in the orthography with vowel lengthening in their pronunciation. This very problem has been reported for Siane (Papua New Guinea), even though the language has no phonemic length contrast (Lucht 1978). Note that if a language has no vowel length opposition and no tone contours, then just the *a/aa* distinction could be used to represent a tone contrast without any diacritics (Larry Hyman, pers. comm. 1998.)

Before concluding this section, it is worth noting that the spelling rules for tone may need to be sensitive to segmental spelling rules, and so the two kinds should not be considered independently. For example, like most Bantoid languages, Dschang manifests vowel elision whereby a word-initial vowel is not pronounced when the previous word ends with a vowel. Nevertheless, the word-initial vowel is always written. What tone should be marked on this silent vowel? Some examples that were found in actual texts are given in Table 4, showing the range of possibilities that speakers came up with. (The transcriptions in Table 4 use the following diacritics: ´=high, `=low, !=downstep.)

Table 4 Tone Marking in the Presence of Elision

Transcription	*Gloss*	*Orthographic Options*
[èkùp ꜜtsó]	*peelings*	ekup ētsó, ekup etsō
[èpáp ꜜpáp]	*spotted*	epáp épāp, epáp ēpáp, epáp epāp
[tèkú ꜜsé]	*bold*	tekú ésē, tekú ēsé, tekú esē

Thus, if the segmental orthography is somewhat abstract, incorporating such devices as silent letters, then the spelling rules governing the writing of tone must be no less abstract.

In this section I have covered six types of spelling system aimed at reducing the number of syllables which have to be marked with a diacritic, or the number of different diacritics. The most promising of these were: (i) to represent the most frequent tone by the absence of any symbol; (ii) to mark tone where it is maximally predictive of word or phrase level tone melodies; (iii) to write contour tones using two vowels, avoiding ǎ/â in languages with no vowel length contrast; and (iv) to collapse tonal distinctions by writing two or more tones using the same symbol.

4 Marking Lexical Tone

The function of tone contrasts can usually be classified as either lexical or grammatical. This point will be illustrated using data from Dschang. The Dschang tone system has many tonal minimal pairs. Example (2) gives transcriptions to illustrate a four-way tone contrast. (The degree symbol represents a non-falling final low tone.)

(2)
- a. lètóŋ [–∕] feather
- b. lèꜜtóŋ [–¯] read (n.)
- c. lètòŋ° [––] navel
- d. lètòŋ [–∖] finish (n.)

In this example, tone functions to distinguish lexical items. However, tone also has grammatical uses, as illustrated by the sentences in (3), using the words efɔ *chief*, tɔŋɔ *to call*, and mɔ *child*.

(3)
- a. èfɔ̌ tɔ̀ŋɔ̀ ꜜmɔ́ [–∕––¯] *The chief called a child* (immediate past)
- b. èfɔ̀ tɔ̀ŋɔ̀ ꜜmɔ́ [––––¯] *The chief calls a child* (present)
- c. èfɔ̀ ꜜtɔ̀ŋɔ̀ ꜜmɔ́ [––__–] *The chief will call a child* (immediate future)

Different ways of marking grammatical tone will be treated in §5. Here, we consider a variety of ways of marking lexical tone.

4.1 Marking only minimal pairs

Suppose we make the assumption that a distinction should only be marked when it carries significant functional load. At first glance, and even upon reflection, it is difficult to say how functional load should be measured in the first place. 'So far, the various measures that have been proposed by linguists cannot lay claim to the precision that their mathematical sophistication might appear to invest them with' (Lyons 1971:83). Lyons identifies three reasons for the difficulty. First, the functional load of a contrast is not fixed once and for all, but may vary depending upon position within the word. Second, there is the issue of assessing the frequency of the contrastive words in naturally occurring speech; if no minimal pairs can be found involving high-frequency words then the need to represent the distinction is greatly diminished. Finally, while a contrast may distinguish many minimal pairs, the words concerned may never appear in the same syntactic environment, in which case nothing may be lost by not representing the contrast.

To illustrate the problems with interpreting functional load too literally, I shall give some brief examples from the non-tonal languages English, German and Terena. Consider first the well-known example of word-initial *th* in English. The grammatical words such as the demonstratives *this* and *that* are pronounced with ð. The lexical words such as *think* and *thank* use θ. Now one could argue that the grammatical factor militates in favour of collapsing the two phonemes into one grapheme, as is the case in the English orthography. However, one could equally argue that an orthography which distinguished ð from θ would provide readers with grammatical information that may help them to parse the sentence, and so the contrast ought to be represented in the written form. A language which does just this is German. In capitalising nouns, the German orthography conveys grammatical information directly, giving the reader visual clues to the structure of a sentence.

Terena, an Arawakan language of Brazil, provides an interesting example of a case where phonemic information of low functional load was omitted, and then reinstated after problems arose.

> Terena has a complex stress system employing two different kinds or degrees of stress which seem to carry only a rather light functional load. There are only a limited number of utterances where stress differentiates meaning or clarifies the grammar. After deciding to mark only part of the system, the analysts discovered the long, hard way that *even though it carried a relatively light share of the load, it was the very thing the people depended upon as a linguistic clue*. (Gudschinsky 1973:127, emphasis mine)

The above discussion shows that notions of functional load based on the significance of a phonemic contrast for distinguishing minimal pairs do not touch the core of the issue. We need to recognise that 'not all levels of linguistic structure are necessarily of equal accessibility' (Klima 1972:62). For example, it is possible to imagine contrasts which might be important in understanding the tone system of a language (e.g. floating tones) and which carry a heavy functional load. Nevertheless, such information may not be accessible to speakers (cf. Klem, 1982:20f). Thus, we need to adopt a broader definition. The *functional load* of an orthographic feature is the extent to which users of the orthography rely on that feature in reading and writing the language. The only way to be certain about issues of functional load is through actual experimentation. Anything less is pure speculation.

Of course, the whole issue can be swept under the rug if we are prepared to assume that the job of an orthography is to distinguish in the written form whatever is distinguished in the spoken form. This is usually done in the name of avoiding *ambiguity* (Pike 1947:37f, fn 49), although the further step of disambiguating homophones is usually not taken. Examples of this method are Ewe (Kwa, Togo & Ghana) and Kasem (Gur, Ghana), where lexical words are marked for tone just when it is necessary to distinguish them from other words (Awedoba 1990). Four problems with this approach can be identified. First, the list of minimal pairs that need to be distinguished with tone marks would be under constant revision as new minimal pairs are discovered, 'to the perennial confusion and discouragement of ... new readers who would resent our sanguine habit of ever expanding the number of words on which we desired them to write tone marks' (Longacre 1964:133). Second, as Kutsch Lojenga (1993:15f) points out, this method conflicts with the practice of omitting the most common tone. When a reader encounters a word unmarked for tone, is that because it is the only word with those segments, or because it is a member of a minimal pair and has the unmarked value? Third, if minimal pairs are distinguished only when there is potential for ambiguity in a particular text, the spelling of a word becomes text-dependent, reducing the ease of word identification and hampering spelling standardisation. Finally, Mfonyam (1990:23) reminds us that 'people do not always think contrastively. When we write, we are not always conscious of potential ambiguities'.[8]

This might seem like an open and shut case against the marking of tone just where it is necessary to avoid ambiguity. However, each of the above four problems can be addressed. First, if a good lexicon is available, then the only new minimal pairs are likely to involve specialist, low frequency vocabulary items which will normally not be confused with higher frequency words. Second, it is not difficult to sidestep Kutsch-Lojenga's objection by adopting an extra tone mark in place of the zero tone grapheme, so that all members of a minimal set are marked for tone. Third, we can insist that the spelling of words is fixed, and does not depend on the possibility for ambiguity in a particular text, or we might argue that text-dependent spellings do not pose a serious problem. There are two possible responses to the fourth problem. One possibility is to simply require that users of the orthography memorise by rote a list of those words that are marked for tone in order to avoid ambiguity. Another possibility is to contest the assumption behind Mfonyam's remark that writers do not 'think contrastively'. Recall the 4-way minimal set from (2). A Dschang speaker writing a story about a bird might use the word letɔŋ *feather* without tone marks, since none of the other three words (*read*, *navel* or *finish*) come to mind at the time. But surely these other words will not come to the mind of the reader either; like writers, readers are not 'thinking contrastively'. If there are high frequency minimal pairs in close semantic association, they will come to the mind of writers, or else writers will learn from experience which words to mark for tone. Thus with due attention, the four problems listed above may be circumvented.

[8]This observation has also been made by other scholars, eg. Wiesemann et al. (1983:157). I have not been able to identify earlier sources.

There is a non-diacritic approach to distinguishing minimal pairs which moves still further away from phonemic orthography. 'It is possible to err on the side of leaving out or eliminating contrasts which are crucial to the correct reading and understanding of written materials; but if only a few key contrasts are lost, they can sometimes be restored by making arbitrary spelling distinctions between the resulting pairs of homonymous words' (Powlison 1968:80). A simple example of this is in Etung, to be discussed in more detail in §5, where á *they* is written as 'ah' while à *he* is written as 'a'. Thus the purely tonal distinction is represented, though without using tone diacritics. Another example of this practice is in Gimira-Bench, an Omotic language of Ethiopia, where certain tonal minimal pairs are distinguished by marking one member of the pair with the ḥäwt letter from the Ethiopic script, a letter which is not otherwise required for the language (Mary Breeze, pers. comm. 1997). For a more thoroughgoing example, it is instructive to examine the approach taken by the designers of the Mandingo orthography. Mandingo is actually a cluster of Mande languages (including Bambara, Mandinka and Malinké) spoken in Mali, Senegal and the Gambia. The unified Mandingo orthography is based on Bambara. Where a tonal minimal pair exists, rather than mark the tone, the designers found 'a variant within the Mandingo group whose use is sufficiently widespread for it to be introduced as a replacement for one of the two elements of the pair' (Galtier 1981:121). So, for example, we find *ma* spelled as *ma*, *man*, or *maa*, depending on which member of the minimal set is intended. However, Denis Creissels (pers. comm. 1996) reports that this orthography was never adopted because it proved impossible to persuade speakers of the main dialects to accept spellings based on low-prestige dialects.

A basic assumption about standardisation needs to be addressed in this context. In many established orthographies, diacritics are used optionally. The same is occasionally proposed for tone languages (e.g. Izi, §6). However, this approach is sometimes criticised as being unsystematic and difficult to teach (cf. Mfonyam 1989: 326). Nevertheless, optionality is a useful feature of diacritics which should be exploited:

> It's good to have a systematic tone-marking system for the use of expatriate language-learners; but in some languages, it may be possible to leave out some of the redundant tone markers in literacy materials for children learning to read their native languages and/or for adult literacy programs; and in particular languages, it may be possible to leave out many more tone markers where fluent adult readers/writers are involved. That's a nice feature about diacritics: they are relatively easy to treat as 'optional', leaving them off when they're not needed – by contrast with segmental symbols, which seem more 'essential'. (William Bright, pers. comm. 1997)

In the preceding discussion we have been concerned about avoiding ambiguity. But what *kinds* of ambiguity do we need to be concerned with?[9] In particular, what is the smallest piece of textual material that we want to insist on disambiguating? If it is with words in isolation (which is usually how minimal pairs are established), then all members of a minimal set will have to be distinguished. Since this disambiguation is affected solely by phonological information, this could be called *phonological disambiguation*. If instead we are only concerned with the phrase as the smallest unit, then there is a lower chance for ambiguity to arise. Words that are in

[9] See (Unseth & Unseth 1991) for more detailed treatment of how to classify ambiguities.

complementary distribution will not need to be considered. Equally, if we have two verbs as a tonal minimal pair, but which differ in their transitivity, then the syntactic complementation pattern of the verb will identify which member of the minimal pair is intended. Thus, if ambiguity is viewed at the phrase level and not the word level, we can have *syntactic disambiguation*. Finally, we can allow for *semantic disambiguation*. Suppose again that two verbs comprise a tonal minimal pair, but where one requires a human agent while the other a non-human agent. Then the agent will serve to disambiguate,[10] so long as it appears before or not long after the verb. Semantic disambiguation will not only apply at the sentence level. As larger discourse units are taken into account, the overall topic will rule out many potential ambiguities. Suffice it to say then, that there are various disambiguating factors that need to be considered in deciding which words should be orthographically distinguished in order to avoid ambiguity.

4.2 Marking the isolation tone of a word

Intuition tells us that the isolation form of a word is in some sense basic. Forms of a word that have been perturbed by sandhi are more likely to be ambiguous than the isolation form (cf Voorhoeve 1964:131). Thus, we might mark a word for tone, regardless of its context, using the tone on the word as it is said in isolation. Gudschinsky (1973:123f) recommended this approach for marking tone in Mazatec. While this approach might work for some languages it has some potential drawbacks. For a start, a word might have more than one isolation form and these may be tonally distinct. Or lexical contrasts might be lost in isolation forms (e.g. the tonal distinctions on Etung verbs which do not show up in certain forms; see Table 7 on page 16). Second, the isolation form may carry phrase-level prosodic information – such as boundary tones – to make it into a well-formed free-standing utterance (Pierrehumbert & Beckman 1988), yet this irrelevant phrasal information may still have to be represented, since linguistically naïve readers may not be able to distinguish lexical and phrasal tones. Third, the isolation tone might not be distinctive (contra Voorhoeve). For example, suppose we have two monosyllabic words which are segmentally identical, where one word has high tone and the other has downstepped high. These words cannot be distinguished in isolation, and therefore they will be orthographically identical. The same problem arises for words having lexical floating tones which are only detected in context. Finally, in languages with tone sandhi – *i.e.* morphophonemic and phrase level tone alternations – a given morpheme will have a variable appearance. For bound morphemes, like tense and aspect markers, there will be no isolation form to appeal to. Perhaps some other context could be chosen, such as the appearance of the morphemes on a fixed word. However, for many languages these considerations will not apply, and marking tone as it appears in isolation will be intuitively appealing and easily taught. There is a further pedagogical advantage of marking isolation tone in a language which manifests tone sandhi:

[10] Willie Kinnaird (pers. comm. 1996) has pointed out to me that this approach breaks down in tales where animals and inanimate objects are personified.

> Beginning readers often go through a sentence very slowly and use the isolation tone; then when they know what it's saying they whiz through it using the tones that occur in each specific environment. Therefore, the isolation tone would be the one that would be helpful to them. (Buck 1973:14)

4.3 Marking the lexical underlying tone

Using the theoretical machinery of autosegmental phonology, a large set of phrase-level tone melodies can usually be related back to a small stock of underlying tone sequences on individual words. Perhaps this smaller, more abstract set ought to be the foundation for a tone orthography. I am not aware of this having been tried, so I shall give a hypothetical illustration from Dschang. For Dschang, one can analyse noun roots as being lexically *H*, *HL*, *LH* or *L*, although the surface forms come out quite differently (see Hyman (1985:48) for full details.) We could mark this four-way contrast as in Table 5. The first column gives the hypothetical orthography. The second column gives a tone transcription; note that final low tones are low-falling except where indicated by a degree sign, following standard practice. Note also that all forms are nouns and carry the noun prefix *le-*.

Table 5 Marking Lexical Underlying Tone in Dschang

Orthography	*Tone*	*Root*	*Lexical Tone*	*Gloss*	*Rule*
letɔ̆ŋ	[lə̀tɔ̀ŋ˚]	tɔ̀ŋ´	LH	*to reimburse*	ˇ = LH
letɔ̀ŋ	[lə̀tɔ̀ŋ]	tɔ̀ŋ	L	*navel*	` = L
letɔ̂ŋ	[lə̀ꜜtɔ́ŋ]	tɔ́ŋ`	HL	*to call*	ˆ = HL
letɔ́ŋ	[lə̀tɔ̆ŋ]	tɔ́ŋ	H	*feather*	´ = H

The main point to observe in Table 5 is that the abstract lexical tone is represented directly in the orthography. This example shows us what morphemic writing of tone might look like. It has the advantage that the shape of a given morpheme is invariable and the obvious weakness that it places expectations on the learner which will often be unrealistic.

In this section we have considered a variety of ways of signalling lexical tone contrasts. We began with the idea of only marking minimal pairs for tone. The main problem with this is that people are not necessarily aware of all the minimally contrasting words for each word being written. However, this may not be a problem if the list of words carrying a tone mark is small enough to be easily memorised. Next, we considered the merit of marking tone on words as they are pronounced in isolation, without marking any of the tone perturbations which arise in phrasal contexts. Then we went on to look at a method for balancing the need to keep lexical forms distinct while minimising the number of tone diacritics. Each of these options, and their attendant problems, is worth considering as part of the design or revision of an orthography for a language with lexical tone contrasts. Next, we move on to consider the orthographic representation of grammatical tone contrasts.

Up to this point, we have seen a host of different approaches to tone marking. Several of these are summarised in Table 6, for a sentence meaning *A woman and her three children used to cultivate a field.*

Table 6 Summary of Tone Orthography Options (Dschang)

Ref.	*Description*	*Example Phrase*
§2	shallow tone marking	M̀mɔ̀' mènzwí pɔ́ èwɔ̀ŋkhʉ̀ pí métɛ́t é lɛ́ ńjʉ́'ɛ́ ŋ́gɔ̄ŋ tɛ̀'
§2	zero tone marking	Mmɔ' menzwi pɔ ewɔŋkhʉ pi metɛt e lɛ njʉ'ɛ ŋgɔŋ tɛ'
§3	omit most frequent tone	Mmɔ' menzwí pɔ́ ewɔŋkhʉ pí métɛ́t é lɛ́ ńjʉ́'ɛ́ ŋ́gɔ̄ŋ tɛ'
§3	mark most stable tone	M̀mɔ̀' mènzwi pɔ èwɔ̀ŋkhʉ̀ pi metɛt e lɛ njʉ'ɛ ŋgɔŋ tɛ̀'
§3	mark tone when it changes	M̀mɔ' menzwí pɔ èwɔŋkhʉ pí metɛt e lɛ njʉ'ɛ ŋgɔ̄ŋ tɛ̀'
§3	mark first tone only	M̀mɔ' mènzwi pɔ́ èwɔŋkhʉ pí métɛt é lɛ́ ńjʉ'ɛ ŋ́gɔŋ tɛ̀'
§3	conflate mid and high	Mmɔ' menzwí pɔ́ ewɔŋkhʉ pí métɛ́t é lɛ́ ńjʉ́'ɛ́ ŋ́gɔ́ŋ tɛ'
§4.1	only mark minimal pairs	Mmɔ' mènzwí pɔ ewɔŋkhʉ pi metɛt é lɛ njʉ'ɛ ŋ́gɔ̄ŋ tɛ'
§4.2	mark isolation tone	Mmɔ' menzwí pɔ́ ewɔŋkhʉ pí métɛ́t e lɛ́ ńjʉ́'ɛ́ ŋgɔ̄ŋ tɛ'
§4.3	mark underlying tone	Mmɔ' menzwí pɔ́ ewɔ̀ŋkhʉ̀ pí metɛ́t e lɛ njʉ́'ɛ ŋgɔ̀ŋ tɛ̀'

5 Grammatical tone marking

In the last three sections we have focussed on lexical tone. We now turn to *grammatical* tone, where tone melodies are responsible for conveying grammatical information such as verb tense. This raises a fresh set of issues for orthography design. Naturally, we can mark grammatical tone phonemically, but this means that the reader must learn to associate complex constellations of tone marks with a grammatical meaning. However, recall that the task before us is to 'reduce the *language* to writing' (Pike 1947), not just the *phonology* of the language to writing. As Venezky points out, 'considerations of the language which an orthography is to reflect must be directed towards the entire language structure, and not just its phonology, although in most instances the phonology will be given priority over the other features' (Venezky 1970). From this broader perspective of putting the *language* into writing, we might consider signalling morphological or grammatical information directly. But what sort of non-phonological information could we encode with a diacritic? This section surveys some of the possibilities. Note that this is not an especially new idea; Pike's proposal to mark 'intonation types' using various punctuation symbols is an early precursor to this approach (Pike 1947:222). (Prior to that even, we can identify such devices as the question mark in European languages.)

Consider first the orthography of Etung/Ejagham (Ekoid Bantu, Cameroon) used in the 1960s. The juncture between the prefix and the root is employed in the expression of grammatical contrasts. Edmondson (1969) uses a colon for perfect tense, a space for perfective aspect, an apostrophe for hortative, a hyphen for conditional, and simple juxtaposition for nouns. Table 7 illustrates the orthography using examples from (Edmondson 1969) and John Watters (pers. comm. 1995). Orthographic forms appear in the first column and tone transcriptions are given alongside in the second column. The remaining columns give the gloss and the orthographic rule.

Table 7 Old Tone Orthography for Etung (Cameroon)

Orthography	*Tone*	*Gloss*	*Orthographic Rule*
m:fap	ḿfàp	*I have left*	colon
e:me	émè	*we have swallowed*	= perfect
m fap	m̀fâp	*I left*	space
e me	èmê	*I bought*	= perfective
m'fap	ḿfáp	*let me leave*	apostrophe
e'me	émē	*let me buy*	= hortative
m-fap	ḿfāp	*when I leave*	hyphen
e-me	émě	*when I buy*	= conditional
mfap	m̀făp	*ants*	no symbol
eme	èmè	*neck*	= noun

The table is organised into five groups of two rows. Each pair of rows contains items with contrastive (underlying) tone. In some constructions, such as the perfect and the perfective, the underlying tone contrast is neutralised. Observe that the lexical tone contrast is not represented in the orthography (although it would be quite straightforward to add accent marks to indicate lexical tone; see §6).

This style of 'tone' marking is a radical departure from what we saw in earlier sections. No amount of simplification measures applied to the tone transcriptions in the second column of Table 7 would ever produce the items in the first column. Here, the symbols are *iconic* of grammatical information. This approach amounts to the *morphemic* writing of grammatical tone.

Another case when diacritics have been used to signal grammatical categories directly is Kako (Equatorial Bantu; Cameroon). This case is interesting because it uses *accent* marks – instead of the punctuation symbols seen in Etung – in order to convey grammatical information. And yet the accent marks still have a loose connection with their more standard, tonal interpretation, satisfying the explicit restriction in Cameroon's orthography standard (Tadadjeu & Sadembouo 1979) that accent marks must be used for tone. Some examples are given in Table 8, from (Ernst 1996c).

Table 8 Tone Orthography for Kako (Cameroon)

Orthography	*Tone*					*Orthographic Rule*
A tì ɓeŋgwɛ nyɛ na	à	tì	ɓéŋgwɛ̄	nyɛ́	nā	tì = past negative
he did not follow him	*he*	N-PST	*follow*	*him*	NEG	
A tí ɓeŋgwɛ nyɛ na	à	tí	ɓēŋgwɛ̀	nyɛ́	nā	tí = future negative
he will not follow him	*he*	N-FUT	*follow*	*him*	NEG	
A ti ɓéŋgwɛ nyɛ na	à	tí	ɓēŋgwɛ́	nyɛ́	nā	ti = aorist negative
he does not follow him	*he*	N-AOR	*follow*	*him*	NEG	
A tî ɓeŋgwɛ nyɛ na!	à	tí	ɓēŋgwɛ̄	nyɛ́	nā	tî = injunctive negative
he must not follow him!	*he*	N-INJ	*follow*	*him*	NEG	

While many grammatical categories could conceivably have been distinguished in the Kako tone orthography, it is noteworthy that Ernst only chose a small subset. Detailed grammatical knowledge and extensive practical experience were essential prerequisites for making the right choice.

Despite the phonological rationale for the diacritics (see §7.2), they are taught with reference to the meaning. After considerable practical experience, it was found that this system was significantly easier to use than both the original system (which used Wiesemann's method of marking only the stable (low) tone, presented on page 7) and the method of not marking tone at all. Ernst explains how the new system was established:

> Rather than simply marking a subset of the phonemic tones, which is the standard practice in this country, a more practical approach was taken. Readers were presented with texts unmarked for tone and wherever they consistently encountered a problem in reading a certain word, the *function* of tone at that point was identified and a spelling change was introduced. A few well-placed diacritics enabled the necessary distinctions to be made, while avoiding the usual problems associated with the over-marking and under-marking of tone. (Ernst 1996c:6)

Another language with grammatical tone marking is Komo, a Bantu language spoken in the Democratic Republic of Congo (formerly Zaire). I am grateful to Paul Thomas (pers. comm. 1997) for furnishing me with this example. In the Komo orthography, tone is marked phonemically, but only to the left of the word root. As Table 9 shows, this tone marking scheme means that the grammatical function of tone is encoded in the orthography, while lexical tone contrasts are not represented.

Table 9 Tone Orthography for Komo (Congo)

bhòm *to insult*, bhóm *to do surgery*

Orthography	*Tone*	*Gloss*
bebhomi	bèbhòmí	*we insulted it*
bebhomi	bèbhómí	*we did surgery on it*
bĕbhomi	bĕbhòmí	*we insulted them*
bĕbhomi	bĕbhómí	*we did surgery on them*
babhomigi	bàbhòmìgì	*insulters*
babhomigi	bàbhómìgì	*surgeons*
bábhomigi	bábhòmìgì	*they insulted habitually*
bábhomigi	bábhómígí	*they did surgery habitually*

Thomas reports that only 28 lexical minimal pairs were found in a list of over 3000 words, and of these, about a dozen could be confused in context. This approach to tone marking results in a tone density of under 10%.

In common with other Bantu languages, Komo verbs have segmental suffixes to indicate tense, aspect and mood. These suffixes correlate with distinct tone patterns on the prefixes. If tone were not marked at all, then readers would sometimes have to scan the whole verb (up to ten syllables) in order to pronounce the first syllable with the correct tone; this may present difficulties for many readers. The Komo example demonstrates the value of marking tone even where there is sufficient segmental information to express the grammatical contrasts.

These examples from Etung, Kako and Komo illustrate just a few of the options for marking grammatical tone. All three involved careful decisions about which distinctions needed to be conveyed and which could be under-represented. It should be clear from these examples that a detailed understanding of the contrastive function of tone in a language is critical for work on tone orthography.

6 Lexical and Grammatical Marking

For languages which make use of both lexical and grammatical tone, a considerable amount of ingenuity may be required in order to achieve good visual separation between the two functions. For example, lexical tone might be written using an accent symbol on the root morpheme, while the grammatical tone associated with a certain verb construction might be expressed using some other punctuation symbol. A *hybrid system* for tone marking uses independent methods for conveying lexical and grammatical tone. Even though lexical and grammatical tone may be overlaid in speech in a complex interaction, they are visually distinguished.

In Izi, a Kwa language of Nigeria, tone is used to distinguish present versus future tense, and declarative versus interrogative mood. These grammatical functions are marked using the hyphen and question mark, as shown in Table 10. Tone also functions lexically: 'Where there are contrasts in the same text, paragraph, or sentence, the tones are written as the word occurs in isolation, *i.e.* with no perturbation. Often it suffices to mark the word which occurs less frequently in general' (Meier 1983:7). In the transcriptions in Table 10, a double-acute accent ő represents an extra-high tone.

Table 10 Tone Orthography for Izi (Nigeria)

ké *to tie*		kè *to divide*	
o-o-ke *he-PROG-tie*		o-o-ke *he-PROG-divide*	
Orthography	*Tone*	*Gloss*	*Orthographic Rule*
ooké	őòké	*he is tying*	no symbol
ooke	őóꜜké	*he is dividing*	= present
?ooké?	òòké	*is he tying?*	question marks
?ooke?	òóꜜké	*is he dividing?*	= interrogative
oo-ké	őóꜜké	*he will be tying*	hyphen
oo-ke	őꜜőkè	*he will be dividing*	= future
?oo-ké?	òóꜜké	*will he be tying?*	question marks + hyphen
?oo-ke?	òőkè	*will he be dividing?*	= interrogative future

Notice in Table 10 that the spoken form őóꜜké is ambiguous between *he is dividing* and *he will be tying*. The form òóꜜké is also ambiguous. However these ambiguities are resolved in the orthography. Since grammatical tone does not place any diacritic on the root vowel, this position is available to host a tone mark, as illustrated in Table 10. However, such marks are only used when lexical ambiguity is not resolved by context. Foreign words transliterated from English are marked with an acute accent on the stressed vowel, to help potential bilinguals. A variety of other orthographic devices are used to represent grammatical tone, such as comma for relative clause and apostrophe for a negation construction.

More recently, Snider (1992) has suggested this hybrid approach as a general strategy in tone orthography, for languages in which tone bears a significant lexical load *and* a significant grammatical load. He illustrates this with an example from a hypothetical orthography for Chumburung (Kwa, Ghana), reproduced in Table 11. The verb is underlined to indicate the locus of the marks referred to in the orthographic rule (third column).

Table 11 Hypothetical Tone Orthography for Chumburung (Ghana)

Orthography	*Tone*	*Orthographic Rule*
naatɨ ma yɔ mɨ kɷdɔrɔ̀	nááti̗ɨ́ má yɔ́ mɨ́ kɷ́dɔ́-rɔ̀	
(The) cow won't go to my farm	*cow* NEG *go my* farm-LOC	
naatɨ ma 'yɔ mɨ kɷdɔrɔ̀	náátɨ́ má ꜝyɔ́ mɨ́ kɷ́dɔ́-rɔ̀	apostrophe = IMP
Cow, don't go to my farm	*cow* NEG-IMP *go my* farm-LOC	
naatɨ ma bà mɨ kɷdɔrɔ̀	náátɨ́ má bá ꜝmɨ́ kɷ́dɔ́-rɔ̀	grave = low tone
(The) cow won't come to my farm	*cow* NEG *come my* farm-LOC	
naatɨ ma 'bà mɨ kɷdɔrɔ̀	náátɨ́ má bà ꜝmɨ́ kɷ́dɔ́-rɔ̀	apostrophe = IMP,
Cow, don't come to my farm	*cow* NEG-IMP *come my* farm-LOC	grave = low tone

Observe that the verb can carry two kinds of information, namely a tone mark (for lexical tone) and an apostrophe (for grammatical tone). Note that the apostrophe could be replaced by a comma, as in the English glosses. Snider (pers. comm. 1996) has expressed reservations about this method if it leads to distinct interpretations of the same diacritic. Consider the two grave accents in the third phrase of Table 11. The first appears on the verb bà since this is a low tone verb even though it is pronounced with high tone as [bá] in this particular instance. The second grave accent appears on the noun kɷdɔrɔ̀, and is justified in the orthography on the basis of the pronounced form (leaving high tone unmarked). Of course, if it is possible to demonstrate that the lexical tone on nouns and verbs does not carry a heavy functional load after all, then we can simply mark the grammatical information, as already explained in §5.

This concludes the critical review of tone orthographies, covering spelling rules (§3), lexical tone (§4), grammatical tone (§5), and hybrid systems (§6). Which of these schemes can be adopted in a given language will depend heavily on contingent details about the language itself, such as whether tone is lexical or grammatical, or whether the tone system is shallow or deep (a notion to be defined below). It will also depend on what one believes about the nature of the reading process and the psychological claims one attaches to any phonological analysis that has been undertaken. These theoretical considerations are treated in the next section.

7 Theoretical Considerations

A range of theoretical considerations impact the design of a tone orthography. We begin by considering the contribution that theoretical phonology has to make to orthography in the area of tone. Even the most casual reader of the phonology literature will have observed that theories vary greatly in how phonological information is structured and manipulated. We begin by considering the contribution of structuralist phonemic theory in §7.1, and then generative phonology in §7.2. Next, we consider the notions of orthographic depth (§7.3) and phonological maturity (§7.4).

7.1 Phonemic theory

Pike argued that 'tonemes substituted in morphology or syntax or sandhi should be written as pronounced. Failure to observe this principle obscures the functional system of the language, hinders the natives in learning to read, and imposes unnecessary burdens upon the foreign student' (Pike 1948:37). In order to understand the original motivation for marking tone phonemically, I shall discuss these three points in turn and offer my own critique on each.

The first claim is that unless tone is written phonemically the functional (*i.e.* grammatical) system of the language is obscured. However, as we saw earlier, there are many ways of diverging from the phonemic system in order to clarify the functional system of the language. For example, in Izi (Table 10 on page 18), phonemic tone writing would have *obscured* the functional system, whereas the non-phonemic method of using question marks and hyphens was particularly revealing of the functional system. The second reason Pike advances is that the process of learning to read is hindered unless tone is written phonemically. Pike's assertion here is not an objective finding, but merely a restatement of the original assumption that a writing system is fundamentally a transcription of the spoken language, a controversial position (see §7.3). In the presence of pervasive tone sandhi, writing the tone as it is pronounced will not support a sight vocabulary.

The third point Pike makes is that a non-phonemic orthography 'imposes unnecessary burdens upon the foreign student'. This view would seem to put the needs of the expatriate linguist, missionary, or any other non-native speaker alongside those of the indigenous population. Perhaps the foreigner who designed a particular orthography wanted to be able to read it before having a command of the phonology, the grammar, and the lexicon of the language. However, Voorhoeve (1964:129) cautions us: 'although the foreigner may more quickly learn to read a text aloud if the [tone sandhi] perturbations are indicated, it makes identification of the words, and thus comprehension of the text, more difficult'. Chomsky and Halle reinforce this point:

> Except for unpredictable variants (e.g., *man – men*, *buy – bought*), an optimal orthography would have one representation for each lexical entry. Up to ambiguity, then, such a system would maintain a close correspondence between semantic units and orthographic representations. A system of this sort is of little use for one who wishes to produce tolerable speech without knowing the language – for example, an actor reading lines in a language with which he is unfamiliar. For such purposes a phonetic alphabet, or the regularised phonetic representations called 'phonemic'

> in modern linguistics, would be superior. This, however, is not the function of conventional orthographic systems. They are designed for the use of speakers of the language. (Chomsky & Halle 1968:49)

Concerns about the suitability of a phonemic orthography for the needs of the reader were raised at an early stage. Hockett (1947:231f) reported ways of deviating from a purely phonemic orthography, and favoured the use of 'multiple writings for certain phonemic sequences'. This results in morphemes having a less variable appearance. In the case of Latin, for example, Hockett advocated that the pairs in (4a) be written as in (4b), in order to preserve the form of the root morpheme.

(4) a. ars:artis, noks:noktis, urps:urbis, reks:regis, niks:niwis
    b. arts:artis, nokts:noktis, urbs:urbis, regs:regis, nigws:nigwis

This approach works since the sequences *rts*, *kts*, *rbs*, *gs* and *gws* do not occur word-finally. Thus, both *rs* and *rts* are representations of phonemic /rs/. Bloch follows Hockett in advocating the move away from phonemic writing to a 'normalised' orthography.

> A normalised notation, still firmly based on the phonemic analysis but incorporating the most common or the most important morphophonemic relations – especially those that are automatic – and such grammatic features as word boundaries and pitch morphemes, approaches the character of a practical orthography. ... When used for the writing of connected texts it reveals more of the linguistic structure; and those who already know the language find it easier to read. (Bloch 1950:348)

Interestingly, Bloch singles out 'pitch morphemes' for special treatment, and this supports some of the approaches to writing grammatical tone described in §5 and §6. Normalisation, then, represents an early call within phonemic theory for a move towards more abstract, morphophonemic writing.

7.2 Generative phonology

In time, the 'autonomous' phoneme of structuralist phonology was abandoned. Phonological rules produced a systematic phonetic output from a systematic phonemic (or abstract lexical) input, without the mediation of an autonomous phonemic level of representation. This move had some interesting implications for phonologists' views of orthography. Far from having a junkyard orthography, English was now hailed as having a near-optimal morphophonemic orthography (Chomsky & Halle 1968:49). (Similar claims have been made for French orthography (Schane 1968) and Arabic orthography (Firth 1948).) Venezky (1970) points out that the two-letter words of English tend to be grammatical words, while lexical words which could have been spelt with two letters are written using three. This is done by adding either a vowel (as in *doe*, *toe* and *roe*), or a consonant (as in *ebb*, *add*, *odd* and *egg*). Though not deliberate, this tendency has served to make grammatical and lexical homophones orthographically distinct (as in *be/bee*, *by/buy*, *or/oar*). Klima (1972) broadened the Chomsky/Halle position, showing how the SPE model actually admits no less than six ways of approaching orthography, varying in abstractness from phonetic to morphophonemic. According to Klima, the approach selected depends on which level one takes to be the most accessible to the language-knower.

The first approach within generative phonology to make a significant contribution to tone analysis was autosegmental phonology, and it is interesting to consider what contribution autosegmental tone analysis has to make to tone orthography. However, the idea that the prosodic component of a word could be directly encoded in orthography actually predates autosegmental phonology. Powlison (1968:81) pointed to the possibility of having an orthography which took Harris' *long components* or Firth's *prosodies* into account. Powlison concluded that such a system would require too much linguistic awareness on the part of the user. Yet the phenomenal success of autosegmental phonology in generating insights about tone systems is reason enough to reopen the discussion. For example, the decomposition of contour tones into sequences of level tones, made fully explicit by autosegmental phonology, can be carried out in the orthography (see §3). Other innovation of autosegmental phonology, such as *floating tones* and *boundary tones* might conceivably be represented using punctuation symbols. After all, it is possible to view the English question intonation as a tone sequence that includes a boundary tone that is not linked to any syllable (Pierrehumbert 1980). Additionally, the idea that tones form a melody that is independent of the actual number of syllables the melody is realised on, has potentially important implications for orthography which have never been explored (though see the Engenni example below). It is less clear in general, though, how sophisticated methods for structuring information invoking multidimensional, hierarchical structure could be insightfully mapped onto a linear, segmental orthography. Nevertheless, there have been some intriguing proposals which shed light on the relationship between autosegmental phonology and tone orthography; these are reviewed below.

In Engenni (Edoid, Nigeria), a single diacritic denotes a distinct tone *melody* on a whole phrase. The meaning of a particular melody depends on the context of the phrase, so it is not possible to assign each melody to a specific grammatical meaning. It was found that 'a protracted tone pattern extending over more than one word can be marked by a single symbol. The symbol can be taught as signalling a certain tone pattern without any reference being made to meaning' (Thomas 1972:10). Although Thomas does not explicitly countenance variation in the number of syllables that a tone melody is being mapped onto, it would presumably be possible for a particular symbol to, say, denote a HLH melody regardless of the actual number of syllables involved. Unfortunately I have been unable to obtain more detailed information about the Engenni system.

Kako shows another application of this idea. Consider Table 12, where a circumflex corresponds to a HL melody spread across one to four syllables (Ernst 1996a).

This idea, originally proposed by Keith Snider, is a particularly striking application of autosegmental phonology. Interestingly, however, the use of the circumflex is actually taught with reference to meaning and not to the tone melody itself (Ernst 1996b).

Another long-standing feature of autosegmental phonology, the *obligatory contour principle* (Leben 1973), may also have ramifications for orthography. Recall that in §3, a proposal by Williamson (1984:43) was discussed, whereby the tone of a syllable is marked only when it differs from the preceding tone. If stretches of identical tone correspond to a single, spread tone,

Table 12 Tone Orthography for Injunctive in Kako (Cameroon)

Orthography	*Tone*	*Gloss*	*Orthographic Rule*
A kêl!	à kêl	*he does-INJ*	circumflex = falling tone
A wôo!	à wóò	*he listens-INJ*	circumflex = high-low
A sânaŋgwɛ!	à sánàŋgwɛ̀	*he works-INJ*	circumflex = high-low-low
A jêmbinaŋgwɛ!	à jémbìnàŋgwɛ̀	*he sings-INJ*	circumflex = high-low-low-low

then having one mark for each such tone amounts to adopting a version of the obligatory contour principle that applies over entire phrases.

The contribution of autosegmental phonology to tone orthography remains a relatively unexplored area. From what we have seen above, autosegmental phonology may give us ideas of how to represent tone in languages where (i) there are a limited number of basic tone melodies which are most compactly represented as one orthographic symbol per tone melody rather than one symbol per tone; or (ii) tonally distinct words appear the same in isolation, or where there are morphemes manifested solely by perturbation of the tones of nearby words (*i.e.* through the action of floating tones). In both situations we may want to adopt diacritic symbols which are iconic representations of an underlying tone pattern.

Other theoretical paradigms falling within the ambit of generative phonology, such as underspecification phonology (Archangeli 1988) and lexical phonology (Kiparsky 1985), may also have a contribution to make to orthography design. For example, the underspecification of predictable information in phonology could easily be carried over into orthography, perhaps using the orthographic equivalents of *archisegments*. A language with vowel harmony, such as Turkish, could then be written in a way that preserves the visual form of harmonising affixes. The new levels of representation defined in lexical phonology could be used as the basis for an orthography, as recently advocated by Keith Snider (pers. comm. 1997). However, the implications for orthography of these and other, more recent phonological theories await further investigation. In the next section, we turn to an important contribution from the literature on reading theory.

7.3 Orthographic depth

It is well known in English that a (spoken) word can have multiple meanings. So, for example [siːn] can signify the verb *seen* or the noun *scene*. In this case the words are spelled differently, although they need not be. For example, [bæt] can mean either a striking instrument or a flying mammal, and both words are written *bat*. These are cases of *homophony*, and are illustrated in the first panel of Table 13.

The opposite situation is shown in the second panel of Table 13. Here *allomorphy* gives distinct spoken forms for a single word, depending on whether the following word begins with a vowel or a consonant. As before, the difference in pronunciation may or may not be indicated in the orthography. Finally, we have the situation for words like *wind*, where two completely separate words, having different meanings and different pronunciations, are spelled the same (heterophonic homographs). Of course, Table 13 represents a simplification, since the real situation is an intricate web of homonyms and homographs, further complicated by dialect

Table 13 Homophony and Allomorphy in English

Phonology		*Orthography*		*Meaning*
bæt	—	bat	〈	striking instrument
				flying mammal
siːn	〈	seen	—	(verb)
		scene	—	(noun)
ði	〉	the	—	definite
ðə				determiner
ə	—	a	〉	indefinite
ən	—	an		determiner
waɪnd	〉	wind	〈	(verb)
wɪnd				(noun)

variation. In the following discussion, I will not generally draw a clear distinction between the orthographic needs of readers versus writers. Readers and writers will be considered as orthography *users*, and such notions as *fluency* will be applied to both tasks. Also, the notion of a 'word' will be distinguished from written and spoken forms and from meanings. A word will be treated as an abstract mental entity which matches forms and meanings, corresponding to the various rows in Table 13.

The purpose of Table 13 was to demonstrate the mismatch, or lack of one-to-one correspondence, which exists between sounds and meanings in English. And English is not unique in this respect, of course. Probably all languages exhibit the mismatch to some extent, thanks to the ubiquitous nature of allomorphy and homophony. How does orthography fit into this picture? A good way to approach this question is in terms of *orthographic depth*.

Languages differ in the abstractness of their orthography. In some languages, allomorphy is spelled out, and so there tends to be a one-to-one mapping between written and spoken forms and the underlying words (construed as abstract mental entities). This is known as a *shallow* orthography.[11] For example, Serbo-Croatian has a particularly shallow orthography, in which regional variations in pronunciation are reflected in spelling (Katz & Frost 1992:149). At the opposite end of the spectrum are languages which distinguish homophones or conflate allomorphs in the written form, thereby approximating a one-to-one mapping between the written forms and their meanings. This is known as a *deep* orthography. Standard written Hebrew, for example, has a deep orthography. The omission of vowels emphasises 'the morphologically based consonantal invariance in a given family of word roots' (Katz & Frost 1992:150). Orthographic depth can be quantified (Van den Bosch et al. 1994), but our present interest in the dimension is only qualitative.

The depth of the orthography of a language is not an arbitrary choice. 'Shallow orthographies are characteristic of languages in which morphemic relatives have consistent pronunciations' (Katz & Frost 1992:150). In Serbo-Croation, the morphological processes of inflection and derivation usually do not perturb word stems, and so phonemic representations of stems do not

[11] Note that phonemic orthographies are shallow, but that the neutral term 'shallow' is preferable to 'phonemic' since the latter is a loaded term with a long history of different interpretations.

obscure their morphological content, as would be the case for Hebrew. Thus, the notion of depth relates to the language as well as its orthography:

> Orthographic depth depends upon two variables: the depth of the morphophonological representation itself and the degree to which the orthography approximates this representation. If the morphophonological representation is quite close to the phonetic representation, the orthography will, of course, be close as well. The reader needs to know little phonology because there is little phonology to be known. (Liberman et al. 1980:147)

In the following discussion, I shall use the term *deep tone system* to refer to tone languages having deep morphophonological representations (as far as tone is concerned). Here, we see that the depth of the tone system sets a limit to the depth of the tone orthography. In a tone language where the pitch contour of a morpheme does not change with context, the tone orthography is forced to be relatively shallow: 'there is little tonology to be known'. In the extreme case, tone features are much like any other segmental feature, and do not need to be singled out for special treatment. On the other hand, for a deep tone system, where the pitch contour of a morpheme is significantly different in different contexts, we have a choice. Should the tone orthography be shallow or deep? The relative merits of shallow versus deep orthographies, bottom-up versus top-down reading and analytic versus holistic learning styles have been extensively discussed in the literature (Pearson 1984, Singer & Ruddell 1985, Frost & Katz 1992). Therefore, only a brief review is included here.

In favour of a shallow orthography, the following argument could be advanced. Clearly, our task in designing an orthography is to invent a visual stimulus which cues a word in the mind of a reader. So a visual input serves as an index into the 'mental lexicon'. Now native speakers already possess an indexing structure for accessing the mental lexicon from speech. So it would be maximally efficient if the orthography indicates the speech form as directly as possible. A shallow orthography hooks into an existing indexing mechanism (Gough 1972:335f; Katz & Frost 1992:158). This makes reading and writing into low-level decoding and transcription tasks. Words that have not been encountered before can be written even if their meaning is unknown. Overall, it is an ideal approach for beginning users (reader-writers), since they only need to learn a handful of alphabetic symbols and their correspondence to spoken sounds.

On the other side of the debate, it has often been pointed out that speech and writing are entirely different mediums of communication (Goody 1987:263f). For example, we can countenance more ambiguity in the spoken forms since there is much extralinguistic information to draw on, and clarification can be requested if necessary. It would be surprising if an optimal orthography simply recapitulated the phonemic system of the spoken language. Indeed, some have argued that a rigid adherence to phonemic spelling actually *hinders* word recognition and spelling, while 'benefiting no one aside from those linguists who confuse reading with phonemic transcription' (Venezky 1970). Moving towards a deep orthography allows homophones to be be distinguished while words whose pronunciation varies in context can be given a fixed representation. This latter notion has been referred to using the terms 'unity of visual impression' (Nida 1964:25f) and 'fixed word-images' (Voorhoeve 1964:130). Maintaining a fixed word-image supports readers in developing a *sight vocabulary*, a set of frequently occurring words that can be recognised as a single unit without being broken down into their component letters.

This sight vocabulary will include many grammatical words which, if identified rapidly, provide important early clues to the structure of the sentence and facilitate fluent reading (Venezky 1970).[12]

A further merit of deeper orthographies is that they are less vulnerable to dialect differences and language change, which may affect tone. As Klem recounts: 'when I have asked three or four literate Yoruba speakers from different areas what tones are used on a given word, I have often gotten three or four different answers' (Klem 1982:21). Likewise, Barbara Hollenbach (pers. comm. 1997) reports that across Mixtec dialects (Oto-Manguean, Mexico) the tone patterns vary more than the segments, and so a shallow tone orthography will generally apply only to a single town. To maintain the principles on which a shallow orthography is founded, we must allow the orthographic practice to be a fluid transcription method, which may result in different spellings in different places, or in different generations. Standardisation will render the orthography less phonemic for many speakers. This is less of a problem if a deeper orthography is chosen from the start.

The depth of an orthography is linked to the psycholinguistic process of word recognition. Distinct cognitive mechanisms may be involved in processing shallow versus deep orthographies:[13]

> [The orthographic depth hypothesis] states that shallow orthographies are more easily able to support a word recognition process that involves the language's phonology. In contrast, deep orthographies encourage a reader to process printed words by referring to their morphology via the printed word's visual-orthographic structure. (Katz & Frost 1992:150)

This means that in a shallow orthography, a written form can be pronounced without reference to the mental lexicon, while someone reading a deep orthography must perform lexical access before pronouncing the word. Which is better?

7.4 Phonological maturity

The answer to this question depends on the phonological maturity we expect users of the orthography to have, and the fluency (in reading and writing) that we hope they will achieve. *Phonological maturity* is the degree to which a reader/writer has internalised the principles underlying the relationship between written and spoken forms (cf. Liberman et al. 1980:138). An orthography requiring greater phonological maturity is harder to learn but may facilitate greater fluency. 'For the beginner, the orthography is needed as an indicator for the sounds of words (*inter alia*), but

[12] Note that maintaining fixed word-images is not the same as morphemic writing, in which morphemes are given unique and invariable representations. If two polymorphemic words share all the same morphemes except for one, they still count as distinct words. Accordingly, the doctrine of fixed word-images does not require them to appear more similar to each other than to any other words. The idea of giving a consistent representation to grammatical markers (see §5 and §6) goes beyond the notion of fixed word-images to fixed 'morpheme-images'. I am grateful to Keith Snider (pers. comm. 1997) for impressing this point on me.

[13] Note that although the orthographic depth hypothesis has become the orthodox viewpoint, Seidenberg (1992:103) claims that it 'overstates how much orthographies differ in terms of processing and overlooks some deeper generalisations about how word recognition is achieved.'

for the advanced reader, meanings, not sounds, are needed' (Venezky 1970). The same claim has been made for tone languages (Snider 1992:29). Liberman et al. refine this position and reason, cogently I think, that this conflict of interest between beginning and advanced readers only arises in a certain kind of language:

> Languages with deep morphophonological representations appear to put the phonologically immature learner at odds with the more experienced and phonologically more mature reader. An orthography practical for the former may be cumbersome for the latter. ... [But] a shallower orthography may reduce the reader's opportunities for learning more about his language. (Liberman et al. 1980:148)

It is worth stressing, then, that there is no reason to get caught up in the orthographic depth issue for languages with shallow morphophonological representations. (This includes tone languages in which the tone melody on each word is relatively invariable.) In such languages, there is little phonological maturity to be had in the first place, and adopting a shallow orthography will not pit beginning and advanced users against each other.

The amount of phonological maturity we can expect readers to achieve depends on how steep a learning curve the speakers will tolerate and on the pedagogical resources at our disposal. Our expectations will depend less on linguistic analysis and more on sociological factors. In some settings, the average person might have very limited opportunities for study. New readers may not persevere with a deep orthography long enough to gain enough phonological maturity to make sense of it. So the reward of being able to read may not come early enough to justify the effort. A shallow orthography may be preferable here, since the handicap for a few advanced readers might be outweighed by the needs of the wider community. However, if the pedagogical setting is more favourable, we can opt for a deeper orthography that supports mature users and helps them learn about their language.

8 Some Methodological Principles

What general lessons can be drawn from the above discussion in marking tone in African writing systems? It would be naïve and scholastic to provide a detailed linear procedure for determining the optimal tone orthography for any given language. Moreover, it would risk giving a dangerously patronising perspective that the whole process can be implemented by an external linguist with no regard for sociolinguistic factors (Bird & Hedinger 1997).

Instead, I would like to emphasise some methodological principles which may serve as a guide. This list is not intended to be complete, since it does not cover sociolinguistic, pedagogical and technological factors which also play a crucial role in orthography design. Nor does it cover the representation of tone in dictionaries or in texts intended for non-native speakers of the language.

Identify the function and depth of tone. Tone analysis should proceed at least far enough to identify the function of tone in the language, and to discover the depth of the tone system. This will often require wide-ranging and careful description and analysis of the tone system. However, one does not *necessarily* have to engage the full machinery of abstract theoretical phonology. If tone has no function in the language or if the tone system is shallow then no special treatment is required over and above the traditional phonemic one, resulting in zero marking and shallow marking respectively.

Practice selective under-representation. If tone has a function and if the tone system is deep, then we need to identify the extent to which these functions of tone can be partially under-represented due to the presence of phonological, syntactic and semantic disambiguating factors.

Endeavour to maintain fixed word-images while using only accessible information. We should devise various ways of representing the functions of tone using the orthography, endeavouring to maintain a fixed word-image while using only information that is accessible to the people (including beginners) who will use the orthography.

Test the options and analyse errors. We should empirically evaluate any proposed orthography to have an objective measure of its performance for a range of readers and writers, with a wide variety of text genres, and undertake meticulous error analysis to see which words and constructions cause problems in practice (Bird 1998). And we should pay attention to actual (rather than official) practice with a tone orthography by studying what readers and writers to in situations that are not closely self-monitored.

Many other factors deserve the careful consideration of the orthography designer, whether a native speaker of the language or not. How will a tone marking method be taught? What technological restrictions exist for typography? Does an existing orthography index political, national, ethnic, or linguistic identity? These issues are not dealt with here, since the goal has not been to provide a total strategy for the orthography designer. Rather, the above four methodological principles will aid the generation of new candidate orthographies for tone languages. Deciding between the orthography options for any given language will then require attention to a wide range of pedagogical and sociological issues.

9 Conclusion

In proposing that tone be marked in a certain language, the boast is usually made of how a slight change in tone can effect a radical change in meaning, and therefore tone needs to be marked. The threefold decision to adopt *shallow* tone marking, using *phonemic* analysis as the foundation and representing tones with *diacritics*, seldom receives any substantive justification. We simply identify the tonal phonemes using a standard discovery procedure. We decide what symbols to use, whether numbers, accents or punctuation by a trial and error method (Bolli 1978). And we might play around with various ways of assigning sets of tonemes to a single grapheme. From this naïve perspective, the procedure for coming up with a tone orthography seems relatively straightforward. However, the end result in many languages has been far too much tone marking. As Karl Grebe (pers. comm. 1996) told me: ‘a major part of learning to read Nso’ is learning which 10% of tone marks must be heeded, and simply ignoring all the rest’.

Set against this overuse is the underuse of tone marking; the influence of European orthographic tradition has too often led to the wholesale rejection of tone marking. In some cases, only a thoroughly rehearsed text can be read fluently. Yet zero marking is well-entrenched:

> I not only see Sudanese reading or attempting to read their non-tonal orthographies three words forward and two steps back, attempting to puzzle out the context in order to correctly decipher the message, but they usually continue to do the same thing after we put tones on because they continue to try to read the alphabetics only, as in English. As most adults stick to their lifelong habits, our hope is in the children. But the adults make the policies and often block the tone marks before enough children get a chance to prove the benefits. (Richard Watson, pers. comm. 1996)

In this paper I have surveyed and critically evaluated a wide range of existing tone orthographies, and reviewed the contribution of phonology and reading theory to the problem, introducing useful terminology and criteria for refining our notions of what constitutes a good tone orthography. I have deliberately eschewed offering across-the-board solutions, since the languages and their sociolinguistic settings differ so widely. I have avoided distinguishing 'practical' and 'scientific' orthography, since I do not believe any such distinction exists.

The focus of the study has been on the languages of sub-Saharan Africa. However, the problems with tone marking described here are not unique to Africa. For example, in Mexico, Barbara Hollenbach (pers. comm. 1997) reports that many literacy programmes experience difficulties with teaching tone, and people who do read 'are clearly not using the tone marks except when they have to.' Apparently only the most highly trained native speakers have learned to write the tone. It is expected that many of the ideas described here will carry over to tone languages in other parts of the world.

I have compiled a collection of tone marking schemes *not* with the intention of opening the floodgates to wild new tone orthographies, nor with the intention of undermining the value of linguistic analysis as a precursor to orthography design. Rather, I hope that the insights of some clever orthography designers will be brought to bear in other languages, perhaps liberating people from some unhelpful assumptions about tone marking. A critical review of each scheme has helped us to see the limits to sensible exploration with tone orthography and, in a sense, confines us to a 'sandbox' which we can safely explore. And this does not prevent us from venturing further; we have been equipped with some helpful evaluation methods which can be applied to any tone marking scheme.

Acknowledgements

This paper was written while I was on an extended field trip to Cameroon, where I conducted tone research under the UK Economic and Social Research Council (grant R00023 5540), working under the auspices of SIL. I am grateful to Bill Bright, Nancy Haynes, Robert Hedinger, Barbara Hollenbach, Connie Kutsch Lojenga, Jim Roberts, Keith Snider, Dick Watson and several SIL language teams for stimulating discussions on tone orthography over an extended period, and for valuable feedback on an earlier version of this paper. This paper was originally submitted to the *Journal of Applied Linguistics*, where it benefitted from three anonymous reviews. I thank Keith Snider for providing me with the title for the paper. Larry Hyman and Len Katz gave astute critical feedback on the final manuscript. My activities in Cameroon (January 1995 – June 1997) were covered by two research permits with the Ministry of Scientific and Technical Research of the Cameroon government (numbers 047/MINREST/D00/D20 and 139/MINREST/B00/D00/D20/D21).

References

Archangeli, Diana. (1988). Aspects of underspecification theory. *Phonology*, *5*, 183–207.

Awedoba, A. K.. (1990). Kasem tones and orthography. *Journal of West African Languages*, *20*(2), 31–46.

Baker, Philip., Bari, Abdulaay., Dalby, David., Jatta, Sidia., Mann, Michael., & Saeed, John. (1982). *Writing African Languages: The Search for a Uniform Script*. School of Oriental and African Studies.

Bendor-Samuel, John. (Ed.). (1989). *The Niger-Congo Languages*. University Press of America.

Bird, Steven. (1998). When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon. *Language and Speech*.

Bird, Steven. & Hedinger, Robert. (1997). Orthography and identity in Cameroon. Paper presented at the 96th Annual Meeting of the American Anthropological Association, Washington, November 1997.

Bloch, Bernard. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, *26*, 86–125. Also appeared in Joos, Martin (ed) (1958) *Readings in Linguistics*, pp. 329–48.

Bolli, Margrit. (1978). Writing tone with punctuation marks. *Notes on Literacy*, *23*, 16–18. SIL.

Buck, Marjorie J.. (1973). Lesson in reading tone: Santo Domingo Zapoteco. *Notes on Literacy*, *15*, 9–15. SIL.

Chia, Emmanuel N.. & Kimbi, Joseph C.. (1992). *Guide to the Kom Alphabet* (Revised ed.). Cameroon: SIL.

Chomsky, N.. & Halle, M.. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clements, G. N.. (1979). The description of terraced-level tone languages. *Language*, *55*, 536–58.

Djoumessi, Mathias. (1957). *Syllabaire Bamiléké*. Yaoundé: St Paul.

Edmondson, Tom. (1969). Ejagham Orthography. SIL Archives, Yaoundé.

Ernst, Urs. (1996a). Alphabet et Orthographe du mbonjɔɔ (kakɔ-ouest). Unpublished manuscript, SIL Cameroon.

Ernst, Urs. (1996b). Marquage du Ton en Kakɔ. Unpublished manuscript, SIL Cameroon.

Ernst, Urs. (1996c). Tone Orthography in Kakɔ. Unpublished manuscript, SIL Cameroon.

Essien, Udo E.. (1977). To end ambiguity in a tone language. In *Language and Linguistic Problems in Africa: Proceedings of the VII Conference on African Linguistics*, (pp. 155–67). Columbia, SC: Hornbeam Press.

Firth, John Rupert. (1948). Sounds and prosodies. In *Papers in Linguistics 1934–1951*. London: Oxford University Press (1957).

Frost, Ram. & Katz, Leonard. (Eds.). (1992). *Orthography, Phonology, Morphology and Meaning*, volume 94 of *Advances in Psychology*. Amsterdam: North-Holland.

Galtier, Gérard. (1981). Current problems in the transcription of Bambara and Soninke. In *African Languages: Proceedings of the Meeting of Experts on the Transcription and Harmonization of African Languages, Niamey (Niger), 17–21 July 1978*, (pp. 110–46). Paris: UNESCO.

Goody, Jack. (1987). *The Interface between the Written and the Oral*. Studies in Literacy, Family, Culture and the State. Cambridge University Press.

Gough, Philip B.. (1972). One second of reading. In James F.. Kavanagh & Ignatius G.. Mattingly (Eds.), *Language by Ear and by Eye: The Relationships between Speech and Reading* (pp. 331–58). MIT Press.

Gudschinsky, Sarah C.. (1973). *A Manual of Literacy for Preliterate Peoples*. Ukarumpa, PNG: SIL.

Hartell, Rhonda L.. (Ed.). (1993). *Alphabets of Africa*. Dakar: UNESCO/SIL.

Hockett, Charles F.. (1947). Problems of morphemic analysis. *Language*, *23*, 321–43. Also appeared in Joos, Martin (ed) (1958) *Readings in Linguistics*, pp. 229–42.

Hyman, Larry M.. (1985). Word domains and downstep in Bamileke-Dschang. *Phonology Yearbook*, *2*, 45–83.

International African Institute (1930). *Orthographe Pratique des Langues Africaines*. Paris: Institut International des Langues et Civilisations Africaines.

Jenewari, Charles E.W̃.(̇1989). Ijoid. In John. Bendor-Samuel (Ed.), *The Niger-Congo Languages* (pp. 105–118). University Press of America.

Jones, J. Randall. (1996). Tone in the Kom Noun Phrase. Technical report, SIL Cameroon.

Kashoki, Mubanga E.. (1981). Harmonization of African languages: standardization of orthography in Zambia. In *African Languages: Proceedings of the Meeting of Experts on the Transcription and Harmonization of African Languages, Niamey (Niger), 17–21 July 1978*, (pp. 164–75). Paris: UNESCO.

Katz, Leonard. & Frost, Ram. (1992). The reading process is different for different orthographies: the orthographic depth hypothesis. In *Haskins Laboratories Status Report on Speech Research*, volume 111/112 (pp. 147–60). Haskins Laboratories.

Kiparsky, Paul. (1985). Some consequences of lexical phonology. In Colin J.. Ewen & John M.. Anderson (Eds.), *Phonology Yearbook* (pp. 85–138). Cambridge University Press.

Klem, Herbert V.. (1982). *Oral Communication of the Scripture: Insights from African Oral Art*. Pasadena CA: William Carey Library.

Klima, Edward S.. (1972). How alphabets might reflect language. In James F.. Kavanagh & Ignatius G.. Mattingly (Eds.), *Language by Ear and by Eye: The Relationships between Speech and Reading* (pp. 57–80). MIT Press.

Koffi, Ettien N.. (1994). The representation of tones in the orthography. *Notes on Literacy*, *20*(3), 51–9. SIL.

Kutsch Lojenga, Constance. (1986). Some experiences in writing and teaching tone in Africa. *Notes on Literacy, Special Issue No. 1*, 59–65. SIL.

Kutsch Lojenga, Constance. (1993). The writing and reading of tone in Bantu languages. *Notes on Literacy*, *19*(1), 1–19. SIL.

Leben, William R.. (1973). *Suprasegmental Phonology*. PhD thesis, Massachusetts Institute of Technology.

Liberman, Isabelle., Liberman, Alvin M.., Mattingly, Ignatius., & Shankweiler, Donald. (1980). Orthography and the Beginning Reader. In J.. Kavanagh & R.. Venezky (Eds.), *Orthography, Reading and Dyslexia* chapter 10, (pp. 137–53). Baltimore: University Park Press.

Longacre, Robert E.. ([195?] 1964). A tone orthography for Trique. In William. Smalley (Ed.), *Orthography Studies: Articles on New Writing Systems* (pp. 132–7). London: United Bible Societies. Originally appeared in *The Bible Translator*.

Lucht, Ramona. (1978). Siane tone orthography. *Notes on Literacy*, *24*, 25–8. SIL.

Lyons, John. (1971). *Introduction to Theoretical Linguistics*. Cambridge University Press.

Meier, Inge. (1983). Making tone markings acceptable and easy. *Notes on Linguistics*, *27*, 4–10. SIL.

Mfonyam, Joseph. (1990). Tone analysis and tone orthography. *Journal of West African Languages*, *20*(2), 19–30.

Nida, Eugene A.. ([1954] 1964). Practical limitations to a phonemic alphabet. In William. Smalley (Ed.), *Orthography Studies: Articles on New Writing Systems* (pp. 22–30). London: United Bible Societies. Originally appeared in *The Bible Translator* 5, 1954.

Odden, David. (1995). Tone: African Languages. In John A.Ġoldsmith (Ed.), *The Handbook of Phonological Theory*, volume 1 of *Blackwell Handbooks in Linguistics* (pp. 444–75). Blackwell.

Pearson, P. David. (Ed.). (1984). *Handbook of Reading Research*. Longman.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, Massachusetts Institute of Technology.

Pierrehumbert, Janet B.. & Beckman, Mary E.. (1988). *Japanese Tone Structure*, volume 15 of *Linguistic Inquiry Monographs*. MIT Press.

Pike, Kenneth L.. (1947). *Phonemics: A Technique for Reducing Language to Writing*. Ann Arbor: University of Michigan Press.

Pike, Kenneth L.. (1948). *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion.*, volume 4. Ann Arbor: University of Michigan Publications in Linguistics.

Powlison, Paul S.. (1968). Bases for formulating an efficient orthography. *The Bible Translator*, *19*, 74–91.

Sapir, Edward. (1933). The psychological reality of phonemes. *Bibliographic data missing*.

Schane, Sanford A.. (1968). *French Phonology and Morphology*. MIT Press.

Seidenberg, Mark S.(1992). Beyond orthographic depth in reading: equitable division of labor. In Ram. Frost & Leonard. Katz (Eds.), *Orthography, Phonology, Morphology and Meaning* (pp. 85–118). Amsterdam: North-Holland.

Singer, Harry. & Ruddell, Robert B.. (Eds.). (1985). *Theoretical Models and Processes of Reading*. Newark, Delaware: International Reading Association.

Snider, Keith L.. (1992). ‘Grammatical tone’ and orthography. *Notes on Literacy*, *18*(4), 25–30. SIL.

Tadadjeu, Maurice. & Sadembouo, Etienne. (1979). *Alphabet Générale des Langues Camerounaises*. Departement des Languages Africaines et Linguistique, Université de Yaoundé.

Thomas, Elaine. (1972). Marking tone in Engenni. *Notes on Literacy*, *14*, 9–12. SIL.

Tucker, A. N.. (1971). Orthographic systems and conventions in sub-Saharan Africa. In Thomas A.. Sebeok (Ed.), *Current Trends in Linguistics*, volume 7 (pp. 618–53). Mouton.

Ubels, Edward. & Ubels, Virginia. (1986). Alphabet and Orthography Statement for the Karang Language. Technical report, SIL-Cameroon.

Unseth, Pete. & Unseth, Carole. (1991). Analyzing ambiguity in orthographies. *Notes on Literacy*, *65*, 35–52. SIL.

Van den Bosch, A.., Content, A.., Daelemans, W.., & Gelder, B. De. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*, 178–88.

Venezky, Richard L.. (1970). Principles for the design of practical writing systems. *Anthropological Linguistics*, *12*, 256–70.

Voorhoeve, Jan. (1964). Some problems in writing tone. In William. Smalley (Ed.), *Orthography Studies: Articles on New Writing Systems* (pp. 127–31). London: United Bible Societies.

Welmers, William E.. (1973). *African Language Structures*. University of California Press.

Welmers, William E.. (1976). *A Grammar of Vai*. University of California Publications in Linguistics Volume 84. University of California Press.

Wiesemann, Ursula., Sadembouo, Etienne., & Tadadjeu, Maurice. (1983). *Guide pour le Developpement des Systemes d’Ecriture des Langues Africaines*. SIL and University of Yaoundé. Second Edition, 1988.

Williamson, Kay. (1984). *Practical Orthography in Nigeria*. Ibadan: Heinemann Educational Books (Nig.).