

# When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon

Steven Bird  
University of Edinburgh

December 1998

## **Abstract**

Should an alphabetic orthography for a tone language include tone marks? Opinion and practice are divided along three lines: zero marking, phonemic marking and various reduced marking schemes. This paper examines the success of phonemic tone marking for Dschang, a Grassfields Bantu language which uses tone to distinguish lexical items and some grammatical constructions. Participants with a variety of ages and educational backgrounds, and having different levels of exposure to the orthography were tested on location in the Western Province of Cameroon. All but one had attended classes on tone marking. Participants read texts which were marked and unmarked for tone, then added tone marks to the unmarked texts. Analysis shows that tone marking degrades reading fluency and does not help to resolve tonally ambiguous words. Experienced writers attain an accuracy score of 83.5% in adding tone marks to a text, while inexperienced writers score a mere 53%, which is not much better than chance. The experiment raises serious doubts about the suitability of the phonemic method of marking tone for languages having widespread tone sandhi effects, and lends support to the notion that a writing system should have 'fixed word images'. A critical review of other experimental work on African tone orthography lays the groundwork for the experiment, and contributes to the establishment of a uniform experimental paradigm.

Keywords: tone language; orthography design; orthographic depth; reading fluency; Africa

## Introduction

In a tone language, the pitch on an individual syllable can be contrastive, thereby distinguishing two or more lexical items or grammatical categories (such as verb tense). African tone languages are often written using the IPA-based *Africa Script* (International African Institute, 1930), which provides diacritic symbols such as acute accent for high tone and grave accent for low tone. A set of principles whereby tonal distinctions are represented (or under-represented) orthographically is known as a *tone orthography*. Although some tone languages are not written with tone marks, it will often be convenient to think of them as still having a tone orthography, but with complete under-representation: ‘zero tone marking’.

All too often, tone orthographies are established by fiat and defended by anecdote. Whether or not tone is marked, the most frequently cited justifications offered by the designers are either linguistic analysis, or socio-political factors, or an impressionistic evaluation that ‘we tried it and it seemed to work fine’. This article presents objective evidence that an existing tone orthography for an African tone language actually *hinders* fluent reading and writing. A wide range of subjects were tested, covering different ages, educational backgrounds, and levels of exposure to the orthography. Their poor performance on reading and writing tasks involving tone marking challenged expectations and led to the conclusion that phonemic tone marking is not ideally suited to languages with complex tone sandhi.

This does not mean that tone marking should be abandoned in the language. The finding simply highlights the fact that relatively little is known about the reading process for alphabetic orthographies decorated with tone diacritics. I argue that any consideration of the linguistic and socio-political factors influencing orthography design must be complemented with experimental work that provides an objective evaluation of orthography options.

Most of the existing work on experimenting with orthography is for languages with established orthographies, with the aim of discovering more about the reading process (Henderson, 1984; Frost and Katz, 1992). In the present context, however, the intention is to discover what kind of tone marking for a given language best supports efficient reading, writing and comprehension. In the following sections I shall assume that the aim of the experimental work will be to compare two or more orthography options, where each option is evaluated for its support of fluency. Other dimensions of evaluation, such as the linguistic and socio-political factors mentioned earlier, are not treated here. Nor is the issue of teachability, which depends on the pedagogical resources and opportunities in the language area.

This paper is structured as follows. After a brief descriptive introduction to African tone systems, I survey the experimental work on African tone orthography. Next, the experiment is presented. The paper ends with a discussion of some future prospects and a conclusion. An appendix contains the materials and data.

## African Tone Systems

Almost 2000 languages are spoken in sub-Saharan Africa (Grimes, 1996). The Niger-Congo language family is the largest and by far the most important group as far as tone is concerned. This family stretches from Senegal in the west to Kenya in the east and down into South Africa, and includes the important Bantu language family. Comprehensive surveys of Niger-Congo are available (Welmers, 1973; Bendor-Samuel, 1989), and of tone in general (Fromkin, 1978; van der Hulst and Snider, 1993). Detailed phonetic investigations of African tone languages include Connell and Ladd (1990); Laniran (1992); Liberman et al. (1993).

The vast majority of the Niger-Congo languages are *tonal*, i.e. voice pitch on an individual syllable may carry contrastive meaning, either lexical or grammatical. For example, in the Dschang language of Cameroon, *lɔ̃tɔ̃ŋ* has four lexical meanings, depending solely on the tone: ‘feather’ [–/], ‘reading’ [–], ‘navel’ [––], and ‘finishing’ [– \]. Equally, the phrase *ɔ̃fɔ̃ tɔ̃ŋ mɔ̃* ‘the chief calls the child’, has three grammatical meanings, depending solely on the tone: *near past tense* [–/ ––], *present tense* [– – – –], and *near future tense* [– – \_ \_]. The study of these linguistic systems is known as *tonology*. Some key concepts are introduced below; readers requiring an extended introduction are referred to Odden (1995).

Niger-Congo languages have subject-object-verb or subject-verb-object word order and agglutinative morphology. A major feature of the family is the rich grammatical gender system, whereby nouns are morphologically marked for their *noun class*. For an overview, see Williamson (1989). Niger-Congo languages can generally be analysed as having two, three or four level tones. We refer to these tones – and use them in transcriptions – as follows: low (L) *á*, mid (M) *ā*, and high (H) *á*. Further tone levels may be defined variously as an extra low (XL) *ä*, an extra high (XH) *ǎ*, or a second mid tone (M<sub>2</sub>). Occasionally, a sequence of two level tones appears on a single syllable, creating a rising or falling *contour tone*. These we refer to and transcribe as follows: low-high (LH) *ǎ*, high-low (HL) *â*. Certain sequences of tones give rise to tonal modifications known as *sandhi*; the details vary from one language to the next. As an example of sandhi in Dschang, the word *lɔ̃tɔ̃ŋ* feather is pronounced as *lɔ̃tɔ̃ŋ* when it follows a high tone. The word *lɔ̃tɔ̃ŋ* reading is pronounced as *lɔ̃tɔ̃ŋ* when a low tone follows. For more details of the tone sandhi in this language, see Hyman (1985).

The small inventory of tones discussed above often does not cover the range of tone levels found in a language. For example, we get an arbitrary number of tone levels in the nonsense phrase *ɔ̃ŋ sɔ̃ŋ sɔ̃ŋ sɔ̃ŋ ... bird of bird of bird of bird ...* [– – – –]. However, we can retain the small inventory of tones by introducing *downstep* and *upstep* operators. These can be used to account for local effects, whereby a given tone causes a subsequent tone to be lowered or raised, and for global effects, where a tone causes *all* subsequent tones to be lowered or raised. The latter phenomenon is known as *terracing* and is illustrated in the nonsense example above. For analytical treatments of downstep, upstep and terracing see Clements (1979); Hyman (1979). Though it was first discovered for Niger-Congo languages, downstep has found wide applicability in the analysis of languages outside of Africa, even including non-tonal languages like English. Downstep is usually transcribed using a raised exclamation mark before the syllable, so the above sentence may now be transcribed as *ŋ́ŋ s'óŋ s'óŋ s'óŋ ...*

Downstep is not a tone itself, but it usually arises through the interaction of tones, as illustrated in (1). Here the tones are separated from their host syllables using the notation of *autosegmental phonology* (Leben, 1973; Goldsmith, 1976). Example (1a) gives the word for feather in isolation, and (1b) gives the same word as it appears after a high tone.

- (1) a.   l̩   t̩ɲ  
           | / |  
           L   H
- b.           l̩   t̩ɲ  
           /       |  
           H   L   H

The isolation form of feather (1a) consists lexically of a low tone on the  $\downarrow$  noun class prefix and a high tone on the root  $t\bar{o}\eta$ . Phrase finally, a high tone is pronounced as a rising contour tone if there is a low tone immediately to the left. The dashed line in (1a) indicates that the prefix low tonespreads onto the root, creating the LH sequence – a rising tone. When a high tone precedes the word, as in (1b), then the high tone spreads onto the prefix, delinking the low tone. Thisfloating low tone then gives rise to downstep.

By any standard, the tone systems of most Niger-Congo languages are remarkably complex. What are the implications for orthography? In a transcriptional orthography we might write tone just as it is pronounced. This would be a *shallow* orthography, since there would be a transparent relationship between the written form and its pronunciation. An initial problem for many languages would be to cover an essentially arbitrary number of tone levels with a small number of tone marks. A more substantial problem with this approach, for many languages, would be that the resulting orthography might overwhelm the reader with too much irrelevant detail. In a more abstract, *ordeep* orthography, we might preserve the visual form of words, so that the word  $\bar{b}\bar{t}\bar{o}\eta$  *feather* is always spelled this way, and the reader just has to know that it is pronounced differently in different contexts, a process that might become automatic after some practice. A wealth of literature exists concerning orthography and its relationship to the reading process. See Frost and Katz (1992) for a collection of recent work in this area.

Although informal studies of tone orthography are widespread (see Bird 1998b for a survey), objective experimental work on the writing of tone languages is rarely undertaken; this domain of investigation is uncharted territory. The three studies I am aware of are the subject of the next section.

## Experimenting with Tone Orthography

The experimental work on tone languages is rather limited, focusing primarily on the production and perception of tone (Hombert, 1988; Connell and Ladd, 1990; Whalen and Levitt, 1995; Connell and Bird, 1997). I am aware of three formal experiments on reading African tone languages, and these will be described in this section.<sup>1</sup> Each experiment makes an important contribution to our understanding of tone orthography and the design of tone orthography experiments. First, Essien's experiment on three tone orthographies for Efik is discussed, followed with a review of Mfonyam's work on tone marking in Bafut. Finally we consider Bernard et al.'s experiments on Kom.

### Essien 1977: Efik (Nigeria)

Perhaps the earliest formal experimentation with tone orthography is the work of Essien (1977) on Efik, a Benue-Congo language spoken in Nigeria. The grammatical and lexical function of tone

<sup>1</sup>One other experimental study by Badejo (1989) on Bura (Nigeria), which favours partial tone marking, unfortunately does not present enough detail on the methods or results to include in this survey.

in Efik is demonstrated in (2). Unfortunately no morpheme-level glosses are available. Syllables showing tonal contrasts are underlined.

- (2) a. ékpát úbòk ànwàn mì ókpōn      *My wife's arm is big*  
 b. ékpát úbòk ànwàn mì ókpón      *It is my wife's arm which is big*  
 c. èkpàt úbòk ànwàn mì ókpōn      *My wife's handbag is big*  
 d. èkpàt úbòk ànwàn mì ókpón      *It is my wife's handbag which is big*

Essien's experiment involved three tone marking schemes, zero, 'lexical' and 'grammatical'. In the 'lexical' marking scheme, only nouns and verbs were marked, and the marking was based on the isolation form of the word: 'a high tone verb, for instance, would be marked with a high tone even though in the sentence it would be read with a different tone' (Essien, 1977: 158), a method I shall refer to as *isolation tone marking*. In the 'grammatical' marking scheme, tones were marked as pronounced on the phrase, a method that can be described as *surface tone marking*.

Fifteen inexperienced readers participated in the experiment. All were educated Efik speakers living in the United States. They were drilled for about ten minutes with the isolation and surface tone marking schemes. The experimental materials involved fourteen ambiguous sentences, each approximately ten syllables long, from which three groups of four sentences was selected. Each of the three groups was transcribed according to a different tone marking scheme. Participants were first presented with the set of four sentences unmarked for tone. They were instructed to study the four sentences until a sensible reading was found for each, and then to read them one after the other. This exercise was repeated for the set of four sentences with isolation marking, and then for the set with surface marking.

Subjects were timed on how long they studied a set of four sentences before they started reading, a measure that Essien called 'perception time.' The amount of time taken to read the set aloud was also measured, and this period was called 'vocalisation time.' The results are shown in Table 1.

**Table 1** Mean perception and vocalisation times in seconds for three Efik tone orthographies

	<i>perception</i>	<i>vocalisation</i>
<i>zero</i>	12.25	7.25
<i>isolation</i>	9.5	6.5
<i>surface</i>	14.5	8.5

For both perception and vocalisation time, surface marking was the worst and isolation marking was the best. Essien noted that there were differences between speakers; for example, one was fastest with surface marking, while four were fastest with zero marking. Apparently this difference does not correlate with prior experience: 'Subjects with previous exposure to tone marks showed nothing in their performance relatable to this. They did not do better in either time or correctness than the other subjects' (Essien, 1977: 161).

Essien also compared reading accuracy for the isolation and surface marking schemes. The zero marking scheme was not included; sentences with no tone marks were assumed to have been correctly read since they have many correct readings. Essien reported that only 28.8% of sentences with isolation tone marking and 26.6% of sentences with surface tone marking were read correctly. Overall, eight subjects performed better with the lexical marking, while five performed better with surface marking. Two subjects did equally well with both systems. Essien concluded:

The amount of time it takes to read the short sentences used in this experiment ... is remarkable and seems to suggest a serious problem in the early stages of learning to read Efik with tone marks. If the correlation between tone marks and difficulty is valid, we can imagine what a big help a psychologically preferred tone marking convention would be to the beginning Efik reader. The need to validate the conclusions here strongly suggests more study involving a very large group of Efik readers, as well as many tone marking conventions. (Essien, 1977: 153)

Although Essien's results appear to favour isolation tone marking, no firm conclusion is warranted in the absence of tests for statistical significance of the figures presented in Table 1. Additionally, the accuracy measurements show that readers are not performing better than chance at determining the correct reading of a tone-marked sentence, for either orthography. Perhaps they are simply ignoring the tone marks. More fundamentally, it is difficult to extrapolate from this experiment to normal reading by normal readers. First, the exclusive use of phrases which can be tonally ambiguous in up to five ways, the use of phrases in isolation from context, and the use of nonsense phrases is not typical of normal reading material, as Bernard et al. (1997) also note. Second, the reliance on readers who have had only ten minutes exposure to the writing system is probably unrealistic. Readers cannot be expected to control a tone marking system in just ten minutes, nor will they have acquired a sight vocabulary (*i.e.* a set of words that can be recognised without the need to sound them out). Essien justifies the use of novice readers, asserting that we need 'an orthography that is easiest to read from the learner's point of view; not from the adept's point of view' (Essien, 1977: 162). While one must make concessions to beginning readers, the fate of an orthography should not rest solely with people who first saw the orthography only minutes earlier.

A third problem with extrapolating from these findings to normal reading lies in the way that subjects scanned a whole sentence before reading it aloud. One of the complaints often levelled at zero marking is the way it forces people to read ahead silently for contextual clues. There can be little doubt that zero marking fared so well in this experiment precisely because the reading task gave the subject full access to later material for disambiguation.

Despite these methodological problems, Essien's contribution is of major importance in demonstrating that it is possible to evaluate tone orthographies experimentally. We next turn to the experimental work of Mfonyam (1989) which shows some interesting methodological developments.

### **Mfonyam 1989: Bafut (Cameroon)**

Mfonyam's goal was to establish a tone marking system for a practical orthography of Bafut, a Grassfields Bantu language of Cameroon. His experiment involved four potential tone orthographies. From these candidates the best would be selected and implemented. Mfonyam's starting point was the belief that writing ought to represent speech, and therefore surface tone should be marked (Mfonyam, 1989: 315). However, he observed that the surface marking of tone made reading difficult and so he decided to experiment with different ways of reducing the amount of tone marking. The four systems are described below:

- 1. Stable tone marking** Since low tone is the least variable (phonemic) tone in Bafut, only low tone and contours involving low were marked. So diacritics were used for low (à), high-low (â) and low-mid (ǎ), while all other tones were unmarked.
- 2. Basic tone marking** As a result of his tone analysis, Mfonyam took the basic tone of a noun to be its form when followed by a demonstrative pronoun, and the basic tone of a verb to be the imperative form. Tone perturbations were only spelled out when grammatical ambiguity might arise, otherwise the appearance of a word remained fixed.

**3. Minimal marking** In this system, tones were marked only where lexical or grammatical ambiguity might arise.

**4. Surface marking** Here, all syllables were marked for tone, except for noun prefixes which are elided in fluent speech and for what Mfonyam called ‘phonetic tones’ (the downstepped high !H and the raised low ↑L). Tone contours were marked using a sequence of tone diacritics, and low tone was unmarked.

The first and fourth marking schemes used the tone-to-grapheme correspondences listed in Table 2 (Mfonyam, 1989: 316ff, 515ff). A blank entry in the table indicates that the tone was unmarked. The tones used in the table are: low (L), mid (M), high (H), and sequences of these tones.

**Table 2** Tone to Grapheme Mapping for Bafut

Tone:	L	↑L	M	H	!H	LM	HL	!HL	ML	H!H	LML
Grapheme (1):	`	ˆ		ˆ	ˆ	˘	ˆ	ˆ	ˆ	ˆ	˘
Grapheme (4):			-	ˆ		˘˘	ˆ		˘˘	˘˘	˘˘

The four systems are illustrated in (3) using the text provided by Mfonyam. Unfortunately no gloss or translation is available. I have used the complete 164 word text to establish *atone density* statistic – the ratio of tone marks to tone bearing units – for each system.

### (3) Examples of Bafut Tone Marking Schemes

#### Stable (35%)

À kè bə yíjòŋ Àso fè’è mītaa nyuu ñjoo jì. À fè’è mō, m̀b̀əŋ k̀an loo, a ghɛɛ ŋkwɛrɔ  
akɔŋə a mbo mumaà yì ntsirə ta karə lɔɔ njoo jii mə à kè tswɛ nɛ nyuua.

#### Basic (57%)

A kɪ bə yijon Asō fɛ’ē mītaa nyuu njōō jī. A fɛ’ē mō, mbəŋ kán lóó, a ghɛɛ ŋkwɛrɔ  
akōŋə á mbō múmáa yi ntsírɔ tā kárɔ lɔɔ njōō jii má a kɪ tswé nɛ nyúúa.

#### Minimal (4%)

A kɪ bə yijon Aso fɛ’ɛ mītaa nyuu njoo jì. A fɛ’ɛ mə, mbəŋ kan loo, a ghɛɛ ŋkwɛrɔ  
akɔŋə a mbo muma a yi ntsirə tā karə lɔɔ njoo jii mə a kɪ tswɛ nɪ nyuua.

#### Surface (62%)

A kɪ bə yíjòŋ Asó fè’ē mītáá nyúú ñjōo jī. A fɛ’ɛ mō, mbəŋ kán lōō, á ghɛɛ ŋkwɛrɔ  
akóŋə á mbó múmáa yi ntsírɔ tā kárɔ lɔɔ njōō jii má a kɪ tsw’ɛ nɛ nyuúa.

Mfonyam taught these systems to sixteen Bafut speakers with primary school education and whose ages fell in the range 14–30. The participants were divided into four groups, and each group was taught one of the systems. The writing of tone was taught in the context of an intensive two-week course on reading and writing Bafut. The students were drilled in the hearing, reading and writing of tone. A number of keywords were learnt, and ‘whenever a student had difficulty in telling the tone of a word, he was asked to refer to these words which served as key, or tone reference words’ (Mfonyam, 1989: 330). So, for example, the word àk̀k̀k̀úŋ *owl* was iconic for the low-mid-high sequence. Whenever the low-mid-high sequence occurred in a phrase it could be compared with àk̀k̀k̀úŋ to see if the tones – and thence the diacritics – were the same. I shall call this the *keyword method* of learning a tone orthography.

During the course, participants did eight short class exercises to demonstrate how well they had mastered the reading and writing of tone. At the end of the course they were asked to write a half-page text, marking tone as they had been taught. A thirty-word section of each text was used for evaluation, and only the accuracy of tone marking (and not the other alphabetic symbols) was considered. The group with minimal tone marking was omitted from this writing exercise, since minimal tone marking required writers to make subjective judgements about potential ambiguity which were impossible to score (Mfonyam, 1989: 345). Finally, a reading test measured the accuracy of reading the 164 word text. The results are displayed in Table 3 and indicate the ratio of correctly read or written words to the total number of words.

**Table 3** Results for the Bafut Experiment

<i>Tone Orthography</i>	<i>Class Exercises</i>	<i>Writing Test</i>	<i>Reading Test</i>	<i>Overall Result</i>
Stable	72%	63%	90%	73%
Basic	52%	63%	62%	54%
Minimal	56%	—	60%	57%
Surface	42%	35%	57%	43%

Ignoring for a moment the question of statistical significance, observe that stable tone marking appears to be the best system. Interestingly, surface tone marking – where all potential ambiguities are resolved – fares the worst. Even the minimal marking system with only 4% tone density does better than surface marking. Basic tone marking and minimal marking are about the same. Mfonyam concludes that stable tone marking is the best system for Bafut. He then broadens the conclusion to surface tone marking systems in general, of which stable tone marking is just a special case:

Surface tones should be marked rather than underlying tones or basic tones. This means that a tone orthography that marks underlying tones or basic tones would not be efficient. (Mfonyam, 1989: 346)

I believe we can justifiably remain sceptical about these conclusions. The group size was small with only three to four subjects in each. The writing test was based on a particularly small (30 word) sample, and Mfonyam (1989: 347) himself acknowledges that the writing test was ‘not very conclusive.’ No evidence was given to show that the patterns in Table 3 were significant and not simply due to random variation between individuals. Whether or not one accepts the argument for Bafut, the case for surface tone orthography is not settled here, and experiments on the closely related languages Kom and Dschang (discussed later) are uniformly negative about surface marking.

There is a major flaw with stable tone marking which Mfonyam does not address. Because tone stability is determined on a purely phonological basis, it may not represent grammatical information which is manifested solely through tone sandhi. In this situation we might actually want an unstable tone to be *highlighted* rather than downplayed, because of its communicative function.

Notwithstanding these problems, the design of Mfonyam’s experiment is excellent and his model of parallel groups is worth considering for any experiment on multiple tone orthographies. Mfonyam’s work represents an important methodological advance over Essien’s in his use of an extended training period along with disjoint groups of subjects. This ensures that the participants are well-acquainted with the tone orthography and are not confusing different schemes.

### **Bernard, Mbeh and Handwerker 1995, 1997: Kom (Cameroon)**

Like Bafut, Kom is a Grassfields Bantu language of Cameroon. It has an established tone orthography which uses surface tone marking. Only two tone diacritics are used, and their correspondence with



the tones is given in Table 4 (Chia and Kimbi, 1992: 10; Jones, 1996: 4), where XL is an extra-low tone.

**Table 4** Tone to Grapheme Mapping for Kom

<i>Tone:</i>	H	M	L	XL	HM	HL	HXL	LH	MH
<i>Grapheme:</i>			`	`		^	^		

An example text using this tone orthography is given in (4). A free translation of this text is given in the Appendix.

(4) **Example text from Kom:** *Chimpanzee: king of the forest* Chia and Kimbi (1992: 41).

Na bòbè ìlvì nìn lae tum wâyn ηweyn na wù ndu ì chuf ìvis. Wù fì se sì ki sì ko'si akù, ì yeyn dùyn Bò àkù ta wù n-bâñsì, wu nà kfâ'tè na à n-ghi ìvis. Wù lù ì ko' ndù ì nà sì chu'ti nê ìchì ì fikà' ì. Bò àkù ì syìη bèynsì ìwuyn, bif sì ηweyn nâ à n-ghi ghà a? Wâyn nâ wèyn ì sè sì fâyn ì bè ki na bò vzi bè na yì ko' ì là'ì ηweyn na wù gvi ale' afo a yîna. ...

The Kom tone orthography has a tone density of 40% (Jones, 1996: 20), and was founded by the language committee on the principle of minimising the amount of tone marking without introducing an unacceptable level of ambiguity (G. Schultz, pers. comm. 1996).

Bernard, Mbeh, and Handwerker (1995, 1997) have conducted experiments on Kom tone orthography. Their work is an important development for three reasons. First, it was conducted on a much larger scale than any of the preceding work and it gives, for the first time, a high level of confidence that the findings of a tone orthography experiment generalise to the population as a whole. Second, their work establishes a new standard of rigour in the design and analysis of tone orthography experiments. Finally, Bernard et al. include mature readers in their sample, in contrast to the exclusive use of novice readers by Essien and Mfonyam. Participants come from a wide range of educational backgrounds and ages. This diversity in the pool of subjects makes it possible to determine what class of orthography user has the most difficulty with tone marking. It permits the experimenter to observe which problems disappear with more experience, serving as useful input for teaching tone. Furthermore, it can reveal problems which persist through all levels of experience, perhaps pointing to a problem with the orthography itself. We now review the larger of the two experiments (Bernard et al., 1997).

The experiment used thirteen participants who were literate in English. All but one could already read Kom, nine could write Kom, and six had at some time been teachers of Kom. The materials were based on a set of fifty sentences (6–67 words long) randomly selected from a corpus consisting of proverbs and descriptive texts. Each sentence was written with and without tone marks, and the resulting hundred sentences were randomised. Participants were given a practice exercise lasting fifteen minutes on average.

As in Essien's experiment, participants were asked to study a sentence until a sensible reading was found, and then to read it before going on to the next sentence. The perception and vocalisation times (as defined by Essien) were measured, and the reading was judged to be either correct or incorrect.

Bernard et al. found that the presence of tone marks increased perception time by over 50% and vocalisation time by 15%, while having no effect on the likelihood that a sentence was pronounced correctly. They also report that sentence length had an important effect on accuracy. Giving the reader more context (by using longer sentences) increased the chance of the sentence being read correctly. They reach the following conclusion about marking tone in Kom:

[The experiment] shows that marking tone in long, natural sentences in Kom hinders native speakers in reading those sentences. It hinders the reading of sentences silently, as indicated by the increased perception time. It hinders the reading of sentences orally, as indicated by the increased vocalisation time. And it hinders comprehension as indicated by the greatly increased odds of making errors when reading tone-marked sentences aloud compared with reading tone-unmarked sentences. (Bernard et al., 1997)

Bernard et al. are right to be cautious about generalising this finding to other tone languages. They draw attention to the fact that tone languages use tone in different ways, and call for ‘a program of research to test systematically and comparatively the need for and the effects of marking tone in the languages of the world.’ The need for such a program is clear. And it should be complemented with research on different methods of marking tone for the same language, given the wide variety of tone marking methods that exist (Bird, 1998b).

The experimental paradigm adopted by Bernard et al. has some weaknesses. Like Essien, they chose an unnatural reading task. Normal reading aloud involves simultaneous vocal and visual activity, where the gaze location is usually slightly further along in the text than the word being uttered, and where the two locations are separated by a processing window. Languages evidently do not differ significantly in this property (Gray, 1969: 43ff). Orthographic ambiguity can be resolved without silent reading ahead for contextual clues if there is sufficient disambiguating information inside and to the left of the processing window. If we increase the size of the processing window to include the whole text, by allowing subjects to read ahead silently before reading aloud, then the reader’s reliance on the tone marks is greatly reduced. It is hardly surprising that no disambiguation effect was found.

A second weakness in the design of the experiment is the accuracy measure. Essien had no measure for accuracy; he evaluated only speed. Bernard et al. have already made an important improvement in assessing accuracy. However, they use a binary variable: a sentence with five tone errors receives the same score as a sentence with just one. This measure is too coarse to pick up differential *rates* of error across a set of candidate orthographies. And it is too coarse to distinguish between different *kinds* of tone error, such as pronunciation versus comprehension errors.

A third issue, impressed upon me by William Bright (pers. comm. 1998), is that the many-to-one mapping of tones onto graphemes for Kom, explained at the start of this section, may be at fault. It is at least conceivable that the findings of Bernard et al. demonstrate that this mapping does not correspond adequately to the tonological structure of the language, rather than anything more general about the wisdom of tone marking.

This concludes the discussion of experimental work on African tone orthography. Although each of the three experiments uses different methods, different kinds of subjects and different languages, all agree that full surface tone marking is not optimal. The high tone density which results from surface tone marking imposes too great a cognitive load on readers, and they are unable to use the information conveyed by the tone marks effectively. Essien and Mfonyam both reported that a reduced tone marking system, having stable visual forms for words, is preferable. The same point has been made in a broader context in the literature on reading:

Why is narrow phonetic transcription an unlikely orthography? The reason must be that the shapes of words in such a transcription are context-sensitive and thus difficult to recognise. (Notice what happens to /hænd/, *hand*, in [hæntuwlz], *hand tools*, [hæŋ grənejd], *hand grenade*, [hæmpɪkt], *hand picked*, etc.). The reader is therefore forced to process the transcription symbol by symbol, a slow and arduous procedure. In Chinese, on the other hand, though word-boundaries are absent, the form of an

orthographic word is constant, or at least not subject to contextual variation. It is suggested that this is a minimal constraint that all writing systems must meet, so that words can serve as units of transcription. (Mattingly, 1992: 18)

Each experiment has been critically reviewed in order to help develop standard criteria for evaluating orthography experiments. The experiment described in the next section uses a different design again, not because this is necessarily superior, but because I sought to expand the set of experimental paradigms that had been used and to contribute yet another data point to the discussion of tone testing methods. The experiment reported here continues the programme of research started by Essien and further explored by Mfonyam and Bernard et al., with the primary goal of discovering better methods of marking tone in practical orthographies for previously unwritten languages.

## An Experiment with Dschang

Dschang is a Grassfields Bantu language known to its speakers as Yémba. It is spoken by upwards of 350,000 people in the Western Province of Cameroon. It has predominantly subject-verb-object word order and limited nominal and verbal morphology. The phonology of Dschang has been treated by Haynes (1989) and Bird (1998a).

### Background

#### Dschang tone orthography

Dschang has multiple tone levels which may be analysed with some abstractness in terms of high, low, and floating tones. Phrase-final low tones may be level or falling; following standard practice, phrase-final level low tones are transcribed  $L^\circ$ . Downstep is found in Dschang, so after any tone there is the potential of a six-way distinction between  $H$ ,  $!H$ ,  $L^\circ$ ,  $!L^\circ$ ,  $L$  and  $!L$ . A seventh possibility,  $!!H$ , equivalent in relative height to  $L$ , has also been documented (Hyman and Tadadjeu, 1976). The seven-way contrast can be found at the juncture between a verb and its object noun, by varying the lexical tone on each and the tonal tense-aspect marker. The language is typologically unusual in its use of tone (Laver, 1994: 472), and few other tone languages have such a complex downstepping system.

Dschang has had a variety of orthographies since the 1920s (Momo, 1997). The present tone orthography was adopted in the mid-1980s. Some text in the orthography is shown in (5).

- (5) Kaŋ pɔ mbhū é lelá' ŋgɔ́ mészō, mbú nziŋé ta' enɔ. Pɔ́ lelá' ínāŋ tɛ eshū' amɔ́ álí'í, mbé á ápa, ndək ŋgū́ á ŋkā' ŋiŋ njuú a apumā.

This is a surface tone orthography, where the written form is based on pronunciation in context. The tone density of the orthography is exceptionally high at 58%. High tone is marked with acute accent while mid tone (*i.e.* downstepped high), is written using the macron (overbar) symbol. Both low and low-falling tone are written the same, using the zero mark.

### Objectives

The main aim of the experiment was to evaluate the existing tone orthography of Dschang. The first objective was to verify the assumption that the orthography with tone marks was an improvement on the orthography without tone marks. The second was to learn whether beginning or experienced readers were better served by the tone marks. Beginning readers, with slower reading rates and

lower comprehension, might be unable to profit from context efficiently and so might rely on the disambiguating function of tone marks. Or mature readers might have learned to capitalise on the information encoded in the tone marks. Therefore the first objective would require the use of both beginner and mature readers. A third objective for reading was to discover words and constructions which are liable to be confused, in either the existing or the zero tone marking system. A fourth objective was to assess subjects' active knowledge of the tone orthography by asking them to add tone marks to an unmarked text. Finally, I wanted to gauge attitudes to tone marking. Do people use tone marks in situations which are not closely monitored, such as personal correspondence? Is their level of self confidence reflected in their performance? These and other questions were devised to probe people's perceptions and practices away from the spotlight of the formal reading and writing tests.

## Method

The experiment was conducted in the town of Dschang and the nearby village of Bafou during a ten day period in April, 1997.

## Subjects

Sixteen native speakers of Dschang participated in the experiment for payment. Subjects were chosen with widely varying ages, levels of formal education and reading abilities. Some had little experience of the writing system, having completed just half of the primer, while others had extensive experience (such as one of the primer's authors). The subjects are summarised in Table 5. The performance evaluation in the last column will be explained later.

**Table 5** Subjects in the Dschang Experiment

<i>Subject Id</i>	<i>Age</i>	<i>Sex</i>	<i>Education</i>	<i>Employment</i>	<i>Cluster</i>
11/BT	53	M	primary	teacher	-
12/CT	65	M	primary	retired agric tech	-
14/PA	25	M	vocational	literacy	very good
21/CZ	20	F	secondary	student	fair
23/PN	59	M	vocational	teacher	good
24/EN	53	M	vocational	teacher	fair
25/MK	51	F	primary	housewife	fair
38/JG	32	M	tertiary	translator	very good
41/HN	43	F	primary	seamstress	poor
42/VZ	25	F	primary	housewife	-
44/MD	39	F	primary	housewife	-
46/TK	25	M	primary	unemployed	poor
47/ET	30	M	tertiary	literacy	good
48/AT	32	M	secondary	literacy	very good
49/RD	23	M	primary	unemployed	-
50/BM	23	F	primary	literacy (part-time)	poor

Participants were classed as having a low reading ability if they were unable to read more than one word at a time. The five people in this category (having a dash in the cluster column of Table 5) were discarded since this style of reading produces the isolation tone of words and our present concern is the tonally correct pronunciation of phrases. Of the remaining participants, all but one (41/HN)

had attended literacy classes using a transitional primer designed for people literate in French (Harro et al., 1990). In this primer, the first three lessons teach tone awareness using minimal pairs, along with keywords for memorising tone marks.

### Materials

Twenty narrative texts were collected having an average length of 200 words. From these, four were selected which were of similar style, length and difficulty. (These texts are included in the Appendix.) The texts were chosen so as to not overlap significantly in lexical content. A native speaker I had trained over the preceding two years then prepared each text in two versions for full and zero tone marking,  $F_1 \dots F_4$ ,  $Z_1 \dots Z_4$ , and these were checked by an experienced literacy worker for any errors, especially tone marking errors. Once the corrections were made and double checked, two booklets were prepared, each having four texts, half with full marking and half with zero marking. One booklet contained  $Z_1, Z_2, F_3, F_4$ , while the other contained  $Z_3, Z_4, F_1, F_2$ . Thus both booklets contained the same texts, but the booklets differed in terms of which texts were marked for tone. The booklets had a page of instructions that were written in French since all subjects were also literate in French.<sup>2</sup>

The booklets finished with a one page written questionnaire (also in French) to assess attitudes to tone marking and to elicit a self-assessment of tone-marking ability. This questionnaire was placed at the end of the experiment since, by this time, readers would have been alerted to the problems they have with reading. The questions and responses were in French, and are translated into English here. The questions are listed in (6).

#### (6) Questionnaire on the Dschang tone orthography.

- a. Do you write to your friends in the Dschang language?
- b. In writing personal letters, do you mark tone?
- c. In your letters, how thoroughly do you mark tone?
- d. How do you feel about marking tone on a text, if the text is to be corrected afterwards by a teacher?
- e. What do you advise concerning the marking of tone?
- f. Do you have any other comments?

The focus in the first three questions was on personal correspondence, since the use of tone marking in informal writing was thought to be a good way of judging the overall communicative efficiency of tone marking. The final three questions were intended to elicit the participant's subjective evaluation of the tone orthography.

### Procedure

Subjects were instructed to read the four texts aloud and then to add tone marks to the two unmarked texts. Each subject was tested individually by the same native speaker. This eliminated the possibility that subjects might feel uncomfortable in the presence of a foreigner. Texts were presented one at a time and subjects were not permitted to preview a text before starting to read it. Each of the four texts

<sup>2</sup>I was unable to find anyone literate in Dschang but not literate in French, although I was informed that such people do exist.

was recorded, making a total of 44 recorded texts. Next, subjects were given 20 minutes to add tone marks to the unmarked texts (2.7 seconds per word). A pilot experiment had shown that self-paced marking of tone was problematic as an exercise. Subjects often took as long as possible to mark tone as accurately as possible, sometimes writing tone as slowly as twelve words per minute. For some subjects, progress was so slow that they stopped before finishing because of fatigue. From the pilot experiment it was possible to determine that 20 minutes would be ample for about half the subjects, while the other half would be under pressure to complete in time. This limit made the task more similar to a realistic writing situation which a practical orthography should support, where the goal is efficient communication rather than perfect rendition. The fact that some subjects did not complete the writing task is not a serious problem for the findings, given the way that tone errors are finely categorised. The overall procedure lasted approximately 45 minutes.

## Results

### Scoring

The recordings were evaluated for speed, fluency and accuracy. Speed was measured by timing each reading and calculating the average number of seconds taken to read 100 words (TIME). The measurement of fluency involved the counting of reading disfluencies, and obtaining an average value for 100 words (DISFL). Each repeat of a syllable, word or phrase was counted. If a part of a phrase was repeated three times this was scored as three disfluencies. Hesitations were also scored, although the length of the pause, ranging from a fraction of a second to several seconds, was not taken into account.<sup>3</sup> Reading accuracy was measured by counting tonal errors. Reviewing the recordings, a native speaker judged whether each tone error resulted in a different interpretation of the word or of the grammatical construction (e.g. verb tense). Such errors were classified as comprehension errors (COMP). Remaining tone errors were classified as performance errors (PERF). Errors on contiguous syllables were counted just once. A tone error was counted whether or not the reader subsequently produced a corrected reading, because the correction could be attributable to the use of subsequent context for resolving ambiguity. Non-tonal errors, such as the substitution of a different segment or the omission of a word, were not counted. Where an entirely different lexeme was substituted, this was also not counted as an error, even if the tones were different on the substituted word.

For the writing task, online copies of the texts having no tone marks were edited to add in the tone marks used by each subject. A computer program compared each text with the correct version of the text, syllable by syllable. Both texts were scanned in parallel, and confusion matrices for tones and for sequences of two tones were compiled.

The following subject characteristics were recorded: the subject's age (AGE), gender (SEX), level of education (EDUC), degree of exposure to the primer (TRAINING), and their stated self-confidence in reading (CONF), taken from question (6d). Additionally, subjects were asked which of the texts were already familiar to them from oral literature, and the response was recorded using the binary variable FAMILIAR. Since we need to assess the impact of tone marking on the fluency of an individual, it was necessary to control for individual variation. This was done by performing cluster analysis on TIME and DISFL; the resulting clusters are recorded in the final column of Table 5. This factor is entered into the regression analysis using the variable CATEGORY. To accommodate possible differences in the difficulty of individual texts, three binary variables denoting the texts were incorporated: ANKA,

---

<sup>3</sup>A more detailed model of reading disfluencies would have been desirable, perhaps along the lines of what Shriberg (1994) has developed for disfluencies in non-read speech.

ATAK and LEKAN. The fourth text was indicated by setting these three binary variables to zero. The relationship between these names and the individual texts is shown in the Appendix.

## Reading

Multiple regression was performed on the dataset starting with all independent variables. Variables were eliminated in a stepwise fashion, beginning with the highest  $p$  values. Table 6 reports the results of multiple regression for the reading task. The table is divided into two sections, for reading time and disfluencies respectively. There are four columns of numbers. The first column contains the regression coefficients; this tells us the magnitude of the contribution of each variable. In the first section we see that the presence of tone marking adds 7.5 seconds on average to reading time per 100 words, while in the second section we see that tone marking adds 2.7 disfluencies on average per 100 words. The  $p$  values in the next column record the significance of the contribution of each variable. The final column gives the 95% confidence interval for the coefficient. So, for example, the coefficient of  $-3.7$  for CATEGORY means that there is a 95% probability that the population value for this parameter lies in the range  $(-4.9, -2.5)$ .

**Table 6** Multiple regression results for reading experiment

<i>Variable</i>	<i>Coefficient</i>	<i>P</i>	<i>95% Confidence Interval</i>
<b>Reading time per 100 words (seconds) adjusted <math>R^2 = .5358, p = .0001</math></b>			
CONSTANT	87.179	.0001	(75.062, 99.296)
tone	7.534	.0332	(0.632, 14.435)
category	-4.962	.0044	(-8.277, -1.648)
atak	14.848	.0000	(6.878, 22.818)
training	-13.623	.0446	(-26.898, -0.348)
confidence	-4.292	.0557	(-8.693, 0.110)
<b>Reading disfluencies per 100 words adjusted <math>R^2 = .5461, p = .0001</math></b>			
CONSTANT	18.776	.0001	(14.323, 23.229)
tone	2.662	.0537	(-0.0448, 5.370)
category	-3.733	.0001	(-4.922, -2.545)
atak	4.973	.0026	(1.847, 8.100)
confidence	1.952	.0189	(0.339, 3.565)

Most importantly, these tables show that tone marking makes a negative contribution to fluency, significantly in terms of reading speed and almost significantly in disfluencies. The overall ability of the subject, as identified by the cluster analysis step (CATEGORY), is also a significant predictor. Also, we see that the text ATAK turned out to be more difficult than the others, making a significant contribution in both tables. The presence of the TRAINING variable shows that someone who has completed more of the primer is able to read faster. Interestingly, speakers who are more self-confident are likely to read faster but with *more* disfluencies than speakers who are less self-confident. The other variables do not play a significant role in explaining the performance of readers.

Similar regression analysis of tone performance and tone comprehension errors did not produce any statistically significant findings. This points to the need for further experimentation with better measures of comprehension.

Analysis of the individual tone comprehension errors made for both zero and full marking was revealing. A native speaker classified the errors into lexical and grammatical errors, by listening

to the reader's pronunciation and deciding whether a comprehension error was due to a lexical or grammatical misunderstanding. The results of this exercise are displayed in Table 7.

**Table 7** Analysis of raw tone comprehension errors

	<i>Zero marking</i>	<i>Full marking</i>
<i>Lexical errors</i>	0	3
<i>Grammatical errors</i>	18	11

It is quite remarkable that there were no lexical tone errors for the zero tone orthography (although it would be premature to conclude anything from this fact without further experimentation). If the existing full tone orthography was working perfectly, the right hand column would contain only zeros. Close scrutiny of the individual errors revealed that some readers mistakenly guessed the tense for a text in the unmarked orthography which is conveyed by tone alone, or they failed to guess which of *té before* versus *tè until* was intended. The same errors were found in readings of the marked texts, although they were less frequent. Regardless of which angle we view it from, the performance of the tone orthography for the reading task is surprisingly poor. Next we turn to the writing task.

### Results and discussion: writing

For the analysis of writing, subjects were clustered into two groups on the basis of 14 parameters: the proportion of high, mid and low tones from the correct text that were correctly produced by the subject; the proportion of high, mid, and low tones written by the subject that were correct; and the overall score ( $3 + 3 + 1 = 7$  parameters). The process was repeated for word-initial syllables, since these provided particular difficulty for some writers. In comparing these 14 kinds of error, just two clusters emerged, displayed in Table 8. These will be used as the basis for presenting results in this section.

**Table 8** Subjects

<i>Id</i>	<i>Age</i>	<i>Gender</i>	<i>Education</i>	<i>Employment</i>	<i>Cluster</i>
14/PA	25	M	vocational	literacy	experienced
38/JG	32	M	tertiary	translator	experienced
46/TK	25	M	primary	unemployed	experienced
47/ET	30	M	tertiary	literacy	experienced
48/AT	32	M	secondary	literacy	experienced
50/BM	23	F	primary	literacy	experienced
21/CZ	20	F	secondary	student	inexperienced
23/PN	59	M	vocational	teacher	inexperienced
24/EN	53	M	vocational	teacher	inexperienced
25/MK	51	F	primary	housewife	inexperienced
41/HN	43	F	primary	seamstress	inexperienced

Recall that the writing test had a fixed duration of 20 minutes; in analysing the results, we therefore confine ourselves to the writing errors. In the following discussion, the performance errors and the differences between the two groups are analysed in detail.

Table 9 gives confusion matrices for experienced and inexperienced writers. These were obtained by summing the errors for the members of each group. Both matrices have rows and columns for high (*H*), mid (*M*) and low (*O*) tone, the last of these being unmarked. The first row of the experienced



writers matrix should be read as follows: of all *H* tone marks (acute accent) in the original texts, 1434 were correctly written as *H*, 36 were incorrectly written as *M* and 395 were written as *O* (*i.e.* omitted). For this they earn a raw score of 76.9%. An adjusted score shows how much the raw score is an improvement on random tone marking. For example, if the text frequency of high tone was 50% and subjects scored 80% for the writing of high tone, then the adjusted score would be 60%, because a score of 80 is 60% of the way from the baseline of 50 (random marking) to the maximum of 100 (perfect marking). The adjusted score will be of most interest to us here. Note that figures for correctly written tones are underlined.

**Table 9** Confusion matrices and success rates for all tone marking

		<i>Subject group</i>								
		<i>Experienced writers</i>				<i>Inexperienced writers</i>				
<i>Intended</i>	<i>Observed</i>			<i>Success rate (%)</i>		<i>Observed</i>			<i>Success rate (%)</i>	
	<i>H</i>	<i>M</i>	<i>O</i>	<i>Raw</i>	<i>Adjusted</i>	<i>H</i>	<i>M</i>	<i>O</i>	<i>Raw</i>	<i>Adjusted</i>
<i>H</i>	<u>1434</u>	36	395	76.9	<b>58.3</b>	<u>494</u>	59	1068	30.5	<b>-24.9</b>
<i>M</i>	38	<u>350</u>	138	66.5	<b>61.7</b>	86	<u>88</u>	262	20.2	<b>9.4</b>
<i>O</i>	63	22	<u>1709</u>	95.3	<b>91.7</b>	159	83	<u>1356</u>	84.9	<b>73.1</b>
			<i>Mean:</i>	83.5	<b>73.1</b>			<i>Mean:</i>	53.0	<b>22.0</b>

The most surprising fact about this data, I believe, is the low overall accuracy (73.1%) of the experienced writers. This group of six people includes three full time language development workers who probably control the orthography better than any other speakers of the language. Another surprising fact is that inexperienced writers perform worse than chance when they have to write a high tone. The large figures in the two columns for zero tone demonstrate that this may be functioning as a default category; ‘when in doubt write nothing.’

The low raw score for mid tone for both groups in Table 9 may be put down to the relative instability and lower frequency of the mid tone. One way to quantify the orthographic stability of a tone is to compare isolation forms of words with the various contextual forms that exhibit morphotonemic changes, using natural texts. The stability value of a tone is the probability that any syllable bearing that tone in the isolation form of any word also bears the same tone in phrasal context. Accordingly, the texts used in this experiment were re-marked for tone using the isolation form of each word, as found in the dictionary (Bird and Tadadjeu, 1997). Comparison between the two sets of texts (isolation and contextually marked) revealed the following stability values: high tone: 69.5%; mid tone: 57.5%; low tone (zero): 82.3%.

The hypothesis that tonal stability is a factor in explaining the difficulty of marking tone gains further support from Table 10. Here, just the prefixes are considered, as they are the most tonally-variable morphemes.

Interestingly, the first orthography to mark tone in the language – predating the involvement of any professional linguists – did not mark prefixes for tone (Momo, 1955). Furthermore, tone analysis of the language has revealed that prefixes are heavily influenced by the tone of the preceding word. Hyman (1985) has argued cogently that a prefix forms part of the preceding word for the purposes of phrase level tone rules.

If writing accuracy for a given tone is loosely correlated with the stability of that tone, it is worth checking whether inexperienced writers are simply writing tone as it occurs on words in isolation. However, an inspection of the resulting confusion matrices demonstrated that the inexperienced

**Table 10** Confusion Matrices for Tone Marking on Prefixes

		<i>Subject group</i>								
		<i>Experienced writers</i>				<i>Inexperienced writers</i>				
<i>Intended</i>	<i>Observed</i>	<i>Success rate (%)</i>								
				<i>Raw</i>		<i>Adjusted</i>				
	<i>H</i>	<i>M</i>	<i>0</i>							
<i>H</i>	<u>461</u>	8	143	75.3	<b>52.8</b>	<u>127</u>	6	377	24.9	<b>-37.2</b>
<i>M</i>	1	<u>16</u>	13	53.3	<b>52.2</b>	1	<u>5</u>	23	17.2	<b>15.1</b>
<i>0</i>	40	1	<u>599</u>	93.6	<b>87.2</b>	51	14	<u>523</u>	88.9	<b>76.9</b>
			<i>Mean:</i>	83.9	<b>70.0</b>			<i>Mean:</i>	58.1	<b>23.7</b>

writers are not using the isolation form of a word to guide their tone writing. Other ways of looking at the errors have been tried, including various bigram charts to see if certain tone sequences were particularly difficult, but none of them reveal a clear pattern to the behaviour of either class of writers.

One sure conclusion from the study of these errors is that writers display a clear bias to omit tone marks. We can surmise that, when in doubt, writers tend to leave out tone marks rather than risk writing an incorrect mark. The fact that experienced writers do not exceed 75% accuracy is suggestive of a plateau; years of exposure to the orthography do not naturally give rise to accurate usage. Nor does this low accuracy score appear to matter in the day-to-day use of the written language. The fact that inexperienced writers do not exceed 25% accuracy demonstrates that they have acquired no active knowledge about the use of tone marks.

## General discussion

Why do readers do so well when tone marks are omitted? If we count minimal pairs we might conclude that tone has a high functional load in Dschang. However, the low incidence of comprehension errors (reported in Table 7) is evidence to the contrary. The experiment shows that tone marking in Dschang is crucial only for a small number of grammatical constructions and not for the lexicon at all. Examination of comprehension errors on a much larger corpus of text would be necessary to expand the list of genuine lexical or grammatical ambiguities.

Why do readers do so poorly when tone marks are added? It would be surprising if this was a chance property of the readers that were selected, given that such a wide range of abilities were used. Either the tone orthography itself is problematic, or the way the tone orthography is taught is problematic. I contend that both are true.

By any criterion, Dschang must be viewed as having deep morphophonological representations, as defined by Liberman et al. (1980: 148). In other words, the possible tonal forms of a morpheme are sufficiently diverse that morphophonological representations are distant from the surface forms (see Bird 1998b for further discussion of this point). For ease of reference, such languages will be described as having *deep tone systems*. (If, on the other hand, morphemic relatives have a similar tone melody the language will be described as having *ashallow tone system*.) Both possibilities are illustrated in Figure 1.

On the left in Figure 1 we see a word in a shallow tone system, where its contextual forms are close together reflecting their similarity and the fact that the underlying form (symbolised as a filled circle) is not very abstract. On the right we see a word in a deep tone system, where its contextual forms are

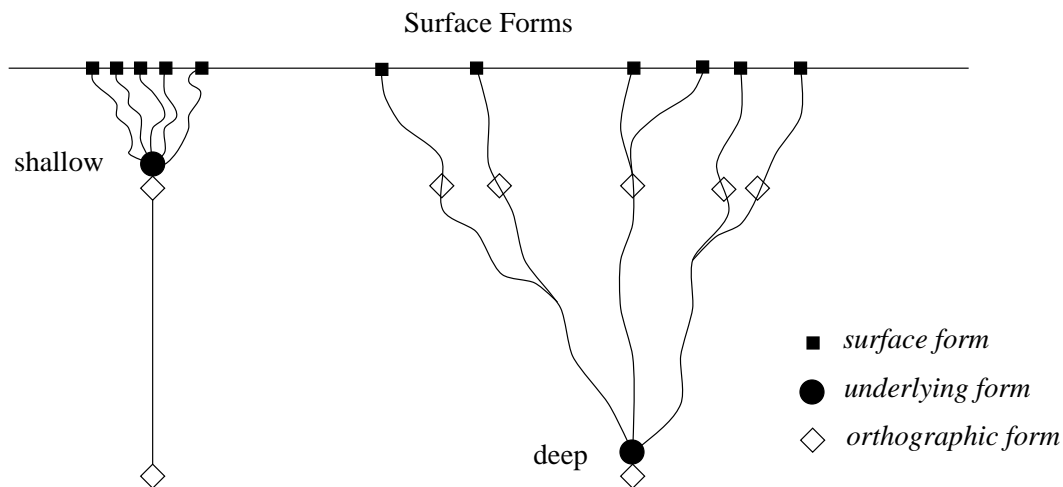


Figure 1: Orthographic and Morphophonological Depth

further apart and the underlying form is much more abstract<sup>4</sup>. The wavy lines in Figure 1 represent derivations. More than one derivation is possible for a word by virtue of its different grammatical and phonological contexts. (For example, the leftmost derivation in both cases might produce the form when the word in question follows another word which ends in a vowel.)

Now, it is also possible to talk about *orthographic* depth. Shallow orthographies tend to spell out allomorphy, as is illustrated in English by the two plural forms *-s* and *-es* (e.g. *cat/cats*, *dish/dishes*). This is symbolised by the row of five diamonds on the right of Figure 1. Each of these denotes a distinct orthographic form. On the left, there is just a single diamond, indicating that the different surface forms are not distinguished by this particular shallow orthography. However, an even shallower orthography could distinguish the five surface forms (possibly by spelling out any allophony). Deep orthographies conflate allomorphs, as is illustrated in English by the single plural form *-s* covering voiced and voiceless allomorphs (e.g. *cat/cats*, *dog/dogs* [dɔgz]). Thus we see a single diamond on the bottom right of Figure 1, corresponding to a unique orthographic form which generalises over all surface forms. Similar abstractness is possible for the tone system on the left, as indicated by the diamond at the bottom left of the figure. However, this abstract orthographic form achieves no additional generality than the shallow form directly above it. Apart from conflating allomorphs in this way, deep orthographies tend to distinguish homophones (e.g. *seen/scene*), a fact that is not represented in the figure.

Observe that in three out of four cases in Figure 1 we have fixed word images. Only in the case of a shallow orthography with a deep tone system do we have multiple forms of words. In general, then, whenever the tone orthography is shallower than the tone system, fixed word images cannot be maintained.

I believe that the current Dschang orthography causes problems for readers precisely because it is a shallow orthography for a deep tone system. This makes it impossible to have the fixed word images which are so important for fluent reading (in particular, reading for comprehension).

<sup>4</sup>This correlation between similarity of surface forms and depth is not accidental, but part of the definition of depth Liberman et al. (1980: 148). Surface forms are more different to the extent that more abstract underlying forms are required to account for them.

A deeper tone orthography would have given a better fit to the tone system, providing better support for fluent reading. I also believe that the existing tone pedagogy for Dschang is poorly matched to the orthography. While the keyword method of teaching tone has had some noteworthy successes elsewhere, these were generally for languages with shallow tone systems. I hypothesise that the keyword method would only work well for Dschang in conjunction with a deeper orthography.

Critical observations were made about the previous experiments on tone orthography, and by the same token, the experiment reported here has its own weaknesses. The original aim to have 40 participants was cut back to 16 when it was more difficult than anticipated to find participants. This figure was further reduced to 11 when five readers could only produce words in isolation. However, even with this small set of participants, behaviour was sufficiently uniform that statistical analysis of the reading data gives good significance values. Another weakness of the current experiment is that it only tested one non-zero tone orthography. I originally hoped to test two tone orthographies, including the new system proposed in (Bird, 1998b). However, it proved impossible to persuade literacy workers to learn a new system, since they did not accept that there were problems with the existing system (Bird and Hedinger, 1997). A further weakness of the experiment, in common with the other experiments, is that it does not have separate exercises that specifically target comprehension. Set against these weaknesses is the fact that the experiment is the first to employ natural exercises on natural materials along with statistical analysis of the results. It is also the first to incorporate detailed error analysis.

### **Results and Discussion for Questionnaire**

The method section above described a short questionnaire and listed the questions in (6). Responses to these questions for all 16 of the original participants are given in the Appendix. It is striking that over half the participants do not mark tone thoroughly on personal correspondence, and that over half would like to see a reduction in the amount of tone marking. So both the behaviour and the wishes of speakers favour reduced tone marking. However, no-one wants zero tone marking. The confidence level in the second last column was actually used in the regression analysis for the reading task; we saw that more confident subjects read faster but made more errors. Evidently, the effective use of the tone orthography is not an area where people have an accurate measure of their own ability. This might be because no evaluation of the acquisition of tone marking skills has been attempted in this language before now. If the teaching programme included tests to identify which areas of the tone orthography students can control, then students and teachers alike would have a more accurate idea of where the strengths and weaknesses lie.

### **The broader context of orthography experiments**

Historically, the linguistic naïveté of the colonists, coupled with the local desire to conform with high-prestige colonial languages has meant the adoption of zero tone orthographies in many countries. On the other side, the chaos created when six European orthographies were let loose in Africa eventually led to the creation of an orthography standard (the Africa Script) which included tone marks. Add to this the desire for orthographies to be visually distinct from the colonial orthographies, as if to underscore the fact that indigenous mother tongues are first class languages in their own right. Further add the conspicuously positivistic faith in the discovery procedures of structuralist phonemic phonology to yield the 'linguistically perfect orthography' (which automatically includes tone marks for tone languages), and we have an unassailable case for phonemic tone marking. It is in this context that the experiment took place.

Both in my survey article (Bird, 1998b) and the design of this experiment, I have not taken sides in the debate between zero and phonemic tone marking. But in proposing alternative orthographies, with whom does the burden of proof lie? Some will point out that the usual criteria of simplicity, parsimony and economy apply. We should not introduce contrasts into an orthography without good cause; each feature of the orthography must pay its own way. Accordingly, we default to zero marking, and experimental evidence purporting to favour a different system must first of all show that it is an improvement on zero marking.

However, others will be quick to point out that an orthography is not a minimal encoding but a transcription system containing plenty of built-in redundancies. In the absence of persuasive arguments to the contrary, an orthography must represent all those contrasts which we know to be linguistically significant. Omitting tone marks for the sake of historical or sociological factors is hardly credible in the face of our ‘modern scientific methods’. The starting point must be the orthography generated by our linguistic technology. Deviations from this can only be sanctioned where there is clear evidence that the linguistically perfect orthography must be tempered by ‘practical considerations’.

I adopt a pragmatic approach. If the intention is to change the status quo, then the status quo must be the starting point, or else the work will not be accessible to its intended audience. For just this reason, some experimental work (e.g. Bernard, Mbeh, and Handwerker 1997) has alienated rather than assuaged the local authorities. Where the national standard or local practice discourages or disallows tone marking (as in many east African countries) zero marking will be our default choice. But this will not be the case for countries having a national standard *requiring* phonemic tone marking (eg. Cameroon). In places where there is no national standard for tone marking (eg. Uganda, Mozambique) the experimenter may be less constrained, though there will normally be a status quo for the language, language group, or region which selects the default tone orthography.

An undesirable consequence of the split described above is that it is difficult to compare experiments on different languages, since they do not tend to test the same range of possibilities. As an illustration of this point, suppose that someone shows that a new tone marking system  $S_1$  is better than zero marking in language  $L_1$ , and someone else demonstrates system  $S_2$  outperforms phonemic marking in language  $L_2$ , what conclusions can be drawn about tone marking across these languages? However, there is a simple solution, which amounts to a recasting of Essien and Mfonyam’s methods which were described at the outset. *Tone orthography experiments should adopt zero and phonemic marking as standard, well-defined controls against which any other tone marking systems are measured.* Now, whether the motivation is to change the status quo or just to do good science, experiments on different languages will be more commensurate.

Finally, note that the default tone orthography does not constrain the process of devising alternative tone orthographies. For example, the language in question may have an existing phonemic tone orthography, and so phonemic tone marking will be the default. However, the other tone orthographies to be tested do not need to take phonemic tone marking as their starting point. It would still be valuable to begin with zero marking and add tone diacritics just where readers experience problems. This point is exemplified in the development of a new tone orthography for Kako (Equatorial Bantu; Cameroon); see Bird (1998b) for a description.

## Future prospects

A new system for marking tone in Dschang was proposed by Maurice Tadadjeu in February 1997 and is currently undergoing a testing phase. Tadadjeu’s system is ingenious in maintaining a fixed form

for roots while allowing the tone marks on prefixes to be changed to reflect grammatical functions. The new system is transparently related to the existing system, and any text can be retranscribed in the new system simply by omitting certain tone marks. A full description of the new system along with experiments will be undertaken in future work. This future work may also test another tone orthography for the language, having tone marks just on a small set of grammatical particles which caused comprehension errors for the unmarked texts (see Table 7).

A separate avenue of exploration would be to study the orthographic depth hypothesis for tone languages. We know that the reading process is different for different orthographies (Katz and Frost, 1992), but how is the reading process affected by different *tone* orthographies? Do we find the same prelexical and postlexical strategies being used with deep and shallow tone orthographies (for the same language) as are reported for the deep and shallow orthographies of Hebrew by Frost (1994)? Some of the interesting properties of Hebrew orthography for experiments on reading (Navon and Shimron, 1984), which stem from the potential for ambiguity through the omission of vowel diacritics, are also to be found in tone orthographies, providing another source of empirical evidence for experiments on reading. Moreover, there is a wealth of experimental paradigms from the reading theory literature (Henderson, 1984; Frost and Katz, 1992) just waiting to be applied to tone orthography. Note that such finely-controlled experiments would shed light on the nature of the reading process for different kinds of tone orthography. This goal needs to be distinguished from the goal of the present experiment, which was to investigate broader issues concerning gross fluency measures, and to make practical judgements about candidate orthographies in the absence of detailed knowledge about the reading process.

An urgent question is whether fixed morpheme images offer something over and above fixed word images. The new Dschang orthography will provide data on this question, but it would also be informative to study some Bantu languages with deep tone systems. For example, consider the data from Chicheŵa (Malawi) in (7) (Kanerva, 1989: 19ff).

- |     |    |      |      |                 |                                |
|-----|----|------|------|-----------------|--------------------------------|
| (7) | a. | múú- | ma-  | leméera         | <i>it is rich (habitual)</i>   |
|     |    | SM-  | TAM- | <i>be-rich</i>  |                                |
|     | b. | múú- | ma-  | leméera         | <i>it is heavy (habitual)</i>  |
|     |    | SM-  | TAM- | <i>be-heavy</i> |                                |
|     | c. | mu-  | ma-  | lémeera         | <i>it was rich (habitual)</i>  |
|     |    | SM-  | TAM- | <i>be-rich</i>  |                                |
|     | d. | mu-  | ma-  | lémééra         | <i>it was heavy (habitual)</i> |
|     |    | SM-  | TAM- | <i>be-heavy</i> |                                |

The tense-aspect marker (TAM) *ma* assigns high tone to its *left* in the present tense (7a,b) and to its *right* in the past tense (7c,d). This would alter the visual forms of the subject marker (SM) and the root respectively. Since the forms in (7) have the status of words, spelling this language with fixed word images would still result in highly variable visual forms of lexical and grammatical morphemes. Lexical distinctions, such as that between *be-rich* and *be-heavy*, are lost in some contexts, as in (7a,b). Would it be desirable to push for fixed morpheme images? I leave this as an open question. We know that morphological awareness helps reading in English (Fowler and Liberman, 1995). And as the following quote attests, spelling similarities between morphemes aid morphological awareness:

We wish to emphasise the fact that the relationship between spelling and morphological knowledge is undoubtedly reciprocal, and that spelling similarities can give rise to morphological insights perhaps at least as readily as morphological awareness can lead to improvements in the skills of writing and reading words in English. (Derwing et al., 1995)

## Conclusion

Should an alphabetic orthography for a tone language include tone marks? I believe this is the wrong question to ask. The range of tone systems, tone orthographies and tone pedagogies is far too great to be addressed by simplistic answers to this question. Instead, we should be asking a different question. Which combination of tone orthography and tone teaching method is best for a given language, taking the language's tone system and sociolinguistic setting into account?

However, the experiment did not succeed in answering even this question, since only one tone orthography and only one tone pedagogy were considered. Nevertheless, I have demonstrated that this combination is actually *worse* than zero tone marking. Reading is slower for marked texts, and the presence of tone marks does not reduce the amount of hesitations and repetitions. No lexical comprehension errors arise from zero marking, even though lexical minimal pairs are the most frequently cited justification for tone marking. The experiment has shown that a shallow tone orthography for a deep tone system, taught using the keyword method, in a sociolinguistic setting which currently provides limited opportunities and motivations for learning the orthography, is a particularly poor combination. I do not wish to generalise further than this, except to note that this experiment counts as another vote in favour of maintaining fixed word images.

This paper began by surveying other experimental work on African tone orthography by Essien (1977), Mfonyam (1989) and Bernard, Mbeh, and Handwerker (1997). While each of these studies used different experimental designs, all agree that full surface tone marking is not optimal for the languages studied. Each study also contributes its own slant on the issue of experimental design in general, and so the critical review serves as a contribution to the development of a standard experimental paradigm. It is in a constructive spirit that I have criticised the experimental work that I hope to encourage.

I complained at the outset that tone orthographies are often established by fiat and defended by anecdote. This applies to any kind of tone orthography, including zero marking. If a tone orthography is unusable, whether through over- or under-representation, or through the wrong kind of tone marking, novice and mature readers alike are faced with a major stumbling block. Where pedagogical opportunities and resources are severely limited, such stumbling blocks can easily have a terminal effect on someone's attempt to learn to read and write. Work on new tone orthographies must not limit itself to a consideration of the linguistic and socio-political factors, important though these are. Rigorous testing of a variety of tone marking options should be a core part of tone orthography design. Such empirical evaluation might even have power of veto over the other dimensions of orthography evaluation, such as linguistic soundness and conformity with existing standards.

Aside from such exhortations, experimenting with tone orthography should be considered because it has many benefits aside from giving objective input to orthography design. An orthography experiment may generate valuable insights about the language under study. A given orthography option may contain unexpected stumbling blocks when a phonological or grammatical distinction is encoded inconsistently, or is not encoded at all (if the expatriate linguist could not hear it), or is encoded but requires information that is inaccessible to the non-linguist user of the orthography. Equally, an orthography option may perform better than expected, challenging our presuppositions. This was the case for zero tone marking in Dschang. Another benefit of experimentation is that it can identify areas of difficulty which should be specifically addressed in pedagogical materials. Experimentation may aid the development of methods for evaluating individuals and identifying problematic categories of learners. If the result of a certain experimental procedure turns out to be a good predictor of the results of other procedures, we have identified a useful diagnostic that may simplify our tests of individual performance. Most importantly, experimentation provides early refutation, so that a bad

design choice will be quashed before it is implemented. And this functions as an insurance policy, affording protection against the chance that someone, someday, will demonstrate that literacy efforts were crippled by an unwise orthography decision.

More experimentation is required. In time, we will have experimental evidence covering a rich typology of tone systems, tone orthographies and tone pedagogies. And this promises to give us a reliable method for deciding which combination of orthography type and teaching method should be tried first in any situation.

## **Acknowledgements**

This paper was written while I was on an extended field trip to Cameroon, where I conducted tone research with the support of the UK Economic and Social Research Council (grant R00023 5540), working under the auspices of SIL Cameroon. In many ways, the ideas contained in this paper have grown out of the SIL experience in Cameroon over the last three decades; they therefore owe a considerable intellectual debt to the organisation. I am grateful to Keith and Mary Beavon, Russ Bernard, Urs Ernst, Robert Hedinger, Randy Jones, Gretchen Harro, Nancy Haynes, Connie Kutsch Lojenga, Dave & Cindy Lux, Joseph Mfonyam, Jim Roberts, George Shultz, Keith Snider and Maurice Tadadjeu for enlightening discussions on tone orthography experimentation. Bill Bright and Larry Hyman gave extensive detailed feedback on the final manuscript, contributing significantly to the readability and coherence of the paper. I am grateful to Pascal-Blaise Kemda, my language associate, for his help in running the experiments and analysing the results, and to Nancy Haynes and Gretchen Harro for providing me with tape recordings of stories in the Dschang language. 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

- B. Rotimi Badejo. An experimental study of tone-marking in Bura. In *Frankfurter Afrikanistische Blätter*, number 1, pages 44–51. Publisher unknown (ISSN 0937-3039), 1989.
- John Bendor-Samuel, editor. *The Niger-Congo Languages*. University Press of America, 1989.
- H. Russell Bernard, George Ngong Mbeh, and W. Penn Handwerker. The tone problem. In A. Traill, R. Vossen, and M. Biesele, editors, *The Complete Linguist — Papers in Memory of Patrick J. Dickens*, pages 27–44. Cologne: Rüdiger Köppe, 1995.
- H. Russell Bernard, George Ngong Mbeh, and W. Penn Handwerker. Does tone need to be marked? Unpublished manuscript, University of Florida, 1997.
- Steven Bird. Dschang syllable structure. In Harry van der Hulst and Nancy Ritter, editors, *The Syllable: Views and Facts*, Studies in Generative Grammar. Mouton-De Gruyter, 1998a. To appear. Originally circulated as Bird (1996), *Dschang Syllable Structure and Moraic Aspiration*, Research Paper EUCCS/RP-69, University of Edinburgh, Centre for Cognitive Science.
- Steven Bird. Strategies for representing tone in african writing systems. *Written Language and Literacy*, 1998b. to appear.
- Steven Bird and Robert Hedinger. Orthography and identity in Cameroon. Paper presented at the 96th Annual Meeting of the American Anthropological Association, Washington, November 1997, 1997.
- Steven Bird and Maurice Tadjadjeu. *Petit Dictionnaire Yémba-Français (Dschang-French Dictionary)*. Cameroon: ANACLAC, 1997.
- Emmanuel N. Chia and Joseph C. Kimbi. *Guide to the Kom Alphabet*. Cameroon: SIL, revised edition, 1992.
- G. N. Clements. The description of terraced-level tone languages. *Language*, 55:536–58, 1979.
- Bruce Connell and Steven Bird. The influence of tone on intrinsic vowel pitch in Mambila and Dschang. In *Proceedings of the Second International Conference on Laryngography*, 1997.
- Bruce Connell and D. Robert Ladd. Aspects of pitch realisation in Yoruba. *Phonology*, 7:1–29, 1990.
- Bruce L. Derwing, Martha L. Smith, and Grace E. Wiebe. On the role of spelling in morpheme recognition: experimental studies with children and adults. In Laurie Beth Feldman, editor, *Morphological Aspects of Language Processing*, pages 3–23. Hillsdale NJ: Lawrence Erlbaum, 1995.
- Udo E. Essien. To end ambiguity in a tone language. In *Language and Linguistic Problems in Africa: Proceedings of the VII Conference on African Linguistics*, pages 155–67. Columbia, SC: Hornbeam Press, 1977.
- Anne E. Fowler and Isabelle Y. Liberman. The role of phonology and orthography in morphological awareness. In Laurie Beth Feldman, editor, *Morphological Aspects of Language Processing*, pages 157–88. Hillsdale NJ: Lawrence Erlbaum, 1995.
- Victoria A. Fromkin, editor. *Tone—A Linguistic Survey*. Academic Press, 1978.
- Ram Frost. Prelexical and postlexical strategies in reading: evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20:116–29, 1994.
- Ram Frost and Leonard Katz, editors. *Orthography, Phonology, Morphology and Meaning*, volume 94 of *Advances in Psychology*. Amsterdam: North-Holland, 1992.
- John A. Goldsmith. *Autosegmental Phonology*. PhD thesis, Massachusetts Institute of Technology, 1976. New York: Garland Publishing. (1979).
- William S. Gray. *The Teaching of Reading and Writing: An International Survey*. Monographs on Fundamental Education. UNESCO, 1969.
- Barbara F. Grimes, editor. *Ethnologue — Languages of the World*. Summer Institute of Linguistics, 13 edition, 1996.
- Gretchen Harro, Nancy Haynes, and Jean-Claude Gnintedem. *Y'ém̄ba, ñpɛ t'í ála'a zímpe: Manuel pour lire et écrire la langue yémba*. Cameroon: SIL, 1990.
- Nancy Haynes. Esquisse phonologique du yemba. In Daniel Barreteau and Robert Hedinger, editors, *Descriptions de Langues Camerounaises*, pages 179–238. Paris: ACCT/ORSTOM, 1989.
- Leslie Henderson, editor. *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology, and Linguistics*, volume 94 of *Advances in psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1984.

- Jean-Marie Hombert. TONPER, un test de perception pour langues tonales applicatoin au Bulu (Sud-Cameroun). In *Pholia*, volume 3, pages 169–81. University of Lyon-II, 1988.
- Larry M. Hyman. A reanalysis of tonal downstep. *Journal of African Languages and Linguistics*, 1:9–29, 1979.
- Larry M. Hyman. Word domains and downstep in Bamileke-Dschang. *Phonology Yearbook*, 2:45–83, 1985.
- Larry M. Hyman and M. Tadadjeu. Floating tones in Mbam-Nkam. In Larry M. Hyman, editor, *Studies in Bantu Tonology*, pages 57–112. University of Southern California, 1976. Occasional Papers in Linguistics, Volume 3.
- International African Institute. *Orthographe Pratique des Langues Africaines*. Paris: Institut International des Langues et Civilisations Africaines, 1930.
- J. Randall Jones. Tone in the Kom noun phrase. Technical report, SIL Cameroon, 1996.
- Jonni Miiikka Kanerva. *Focus and Phrasing in Chicheŵa Phonology*. PhD thesis, Stanford University, 1989.
- Leonard Katz and Ram Frost. The reading process is different for different orthographies: the orthographic depth hypothesis. In *Haskins Laboratories Status Report on Speech Research*, volume 111/112, pages 147–60. Haskins Laboratories, 1992.
- Yetunde Laniran. *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yorùbá*. PhD thesis, Cornell University, 1992.
- John Laver. *Principles of Phonetics*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1994.
- William R. Leben. *Suprasegmental Phonology*. PhD thesis, Massachusetts Institute of Technology, 1973.
- Isabelle Liberman, Alvin M. Liberman, Ignatius Mattingly, and Donald Shankweiler. Orthography and the beginning reader. In J. Kavanagh and R. Venezky, editors, *Orthography, Reading and Dyslexia*, chapter 10, pages 137–53. Baltimore: University Park Press, 1980.
- Mark Liberman, J. Michael Schultz, Soonhyun Hong, and Vincent Okeke. The phonetic interpretation of tone in Igbo. *Phonetica*, 50:147–160, 1993.
- Ignatius G. Mattingly. Linguistic awareness and orthographic form. In Ram Frost and Leonard Katz, editors, *Orthography, Phonology, Morphology and Meaning*, pages 11–26. Amsterdam: North-Holland, 1992.
- Joseph Mfonyam. *Tone in Orthography: The Case of Bafut and Related Languages*. PhD thesis, University of Yaoundé, 1989.
- Gregoire T. N. Momo. *Premiere Vocabulaire Franco-Bamiléké*. Bordeaux: Delmas, 1955.
- Gregoire T. N. Momo. *Le Yemba: Histoire de la Langue Ecrite dans la Menoua*. Dschang: CELY, 1997.
- D. Navon and J. Shimron. Reading Hebrew: how necessary is the graphemic representation of vowels? In Leslie Henderson, editor, *Orthographies and Reading: Perspectives from Cognitive Psychology, Neuropsychology, and Linguistics*, volume 94 of *Advances in psychology*, pages 91–102. Hillsdale, NJ: Lawrence Erlbaum Associates, 1984.
- David Odden. Tone: African languages. In John A. Goldsmith, editor, *The Handbook of Phonological Theory*, Blackwell Handbooks in Linguistics, pages 444–75. Blackwell, 1995.
- Elizabeth Ellen Shriberg. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California at Berkeley, 1994.
- Harry van der Hulst and Keith Snider, editors. *The Phonology of Tone – The Representation of Tonal Register*, volume 17 of *Linguistic models*. Berlin; New York: Mouton de Gruyter, 1993.
- William E. Welmers. *African Language Structures*. University of California Press, 1973.
- D. H. Whalen and Andrea G. Levitt. The universality of intrinsic F<sub>0</sub> of vowels. *Journal of Phonetics*, 23: 349–66, 1995.
- Kay Williamson. Niger-congo overview. In John Bendor-Samuel, editor, *The Niger-Congo Languages*, pages 3–45. University Press of America, 1989.

## Appendix: Materials and Data

### Translation of (4) King of the forest

‘One day, a man sent his son to go out to the neighbourhood and fetch him fire. When he came out of the house and looked up into the forest, his eyes fell on something as red as a glowing fire. It was the exposed anus of the king of the forest (chimpanzee) which he mistook for fire. He hurriedly went near the supposed glowing coal and was trying to break it with a stick. Suddenly the chimpanzee turned around, grabbed him and asked him what he wanted. Scared by the terrifying appearance of the chimpanzee, the child told him that he had been sent by his father to come and invite him to a dinner party. ...’

### Questionnaire responses

Note that people who do not use the language in personal correspondence were still asked the other questions. However, given that there are few other uses for writing, their answers must be regarded as unreliable. Equally, the sensitisation effect of running the experiment at the same time as doing the questionnaire, and the lack of anonymity, are weaknesses in the questionnaire.

**Table 11** Responses to the Questionnaire

<i>Subject Id</i>	<i>Reading ability</i>	<i>a. Write letters?</i>	<i>b. Mark tone?</i>	<i>c. How thoroughly?</i>	<i>d. How confidently?</i>	<i>e. What advice?</i>
11/BT	low	no	always	some words	confident	no change
12/CT	low	no	always	some words	confident	reduce by < 50%
14/PA	high	yes	always	all words	confident	reduce by < 50%
21/CZ	average	no	often	some words	unconfident	no change
23/PN	good	no	always	some words	unconfident	no change
24/EN	average	yes	often	some words	very confident	reduce by < 50%
25/MK	average	no	often	some words	unconfident	reduce by < 50%
38/JG	high	yes	always	all words	confident	reduce by < 50%
41/HN	average	yes	often	some words	confident	reduce by < 50%
42/VZ	low	yes	always	some words	confident	no change
44/MD	low	yes	often	some words	confident	no change
46/TK	average	no	always	all words	very confident	reduce by < 50%
47/ET	good	yes	always	all words	very confident	reduce by < 50%
48/AT	high	yes	always	all words	—	no change
49/RD	low	yes	always	all words	very confident	reduce by < 50%
50/BM	good	yes	always	all words	unconfident	reduce by < 50%

### Answers to the sixth question

**38/JG** I find that tones are always a very difficult chapter in the reading primer.

**46/TK** For myself, I wish that we didn’t mark tone. Unfortunately in our language there are sometimes words with the same spelling but where meaning changes with the tone. If it was possible to write without tone that would be good because teaching where to mark the tones is very difficult.

**47/ET** To reduce the tones, we must not remove in a single sweep what people are already used to marking. This must be done progressively to allow people to re-adjust, which is how people learnt the existing tone marking system.

**48/AT** My advice is not to change the tone marking, since the three tones used currently justify the tonal complexity of the language. In Dschang, only the marking of tone permits certain words to be distinguished. In other words, the sense or the tense changes when the tone changes. In the limit, one can delete less than half the tones so long as that does not cause confusion.

**49/RD** I think the three tones permit good reading, only in a long text, with tone marking the text is over-stuffed.<sup>5</sup> I don't know how one can proceed, since in deleting one of the three tones we could have difficulties.

## Reading materials

### Text 1: The squirrel and the dog (KAN)

This is a traditional story recounted by Maman Alice, recorded by Nancy Haynes and Gretchen Harro, and transcribed by Pascal-Blaise Kemda.

Kaŋ pɔ mbh̄́ é lelá' ɨ́gɔ́ mɛ́sɔ́, mbú n̄́zɨ́ɨ́ ta' enɔ́. Pɔ́ lelá' n̄́nā́ɨ́ tɛ́ esh̄́́ amɔ́' álí'í, mbé á ápa, ndɔ́k ɨ́gɔ́́ á ɨ́kā́' ɨ́ɨ́ njh̄́́ á apumā́. Pɔ́ le gḗ é tṓ á mba, ɨ́gɔ́́ mbɔ́ɨ́. Pɔ́ le gɔ́́ tɛ́ ɨ́kó́ éwú, kaŋ á le mé mbh̄́́ ɨ́gɛ́: « Esó, pá' meɨ́ ɨ́kɔ́́k mbɨ́ɨ́ nzéɨ́ n̄́zā́ɨ́né lā́, meɨ́ ḗ kɔ́' á áth̄́́, é kā́p, o gɔ́́ á é́sɨ́ n̄́nɔ́ɨ́ mbé́tɛ́ n̄́néɨ́ á n̄́tḗ á ápa » . A le lé pū́p pá'ā́, ɨ́kɔ́́' mé ath̄́́, n̄́zɨ́ lékā́p. A le káp̄́, á pé, ɨ́káp̄́, á pé, tɛ́ apa á cɛ́'ɛ́ ɨ́gɔ́́ ndɔ́́, kaŋ á so'ne amɔ́' átā́, n̄́tém á áth̄́́ mbh̄́́, á gɛ́ : ɨ́wā́ɨ́ɨ́, ɨ́wā́ɨ́ɨ́, ɨ́wā́ɨ́ɨ́. Megā́ɨ́ ɨ́ká' é tshɨ́ njú', n̄́zɨ́ lénā́, ɨ́gɛ́: Metsóɨ́ epɨ́ á ɨ́ká'. Kaŋ á le esó'ó á áth̄́́ ndé mé mbh̄́́ ɨ́gɛ́: « Mé ɨ́gɔ́́́ njh̄́́́ n̄́zá́ a lḗyhí? » . ɨ́gɔ́́ mbh̄́́, n̄́néɨ́ á áth̄́́ zɨ́ á lésa mḗth̄́́, ɨ́kɔ́́ ɨ́gɔ́́́. Megā́ɨ́ ɨ́ká' é le esh̄́́', ɨ́gɔ́́ mbh̄́́, n̄́tswá́ á yí tɛ́ á la' ɨ́kwh̄́́. Mé piɨ́ nzɔ́kó́ ánu eɨ́mɔ́ mé́, ndé ɨ́gɛ́: Etsóɨ́ é tɛ́ pɔ́ɨ́.

'Once upon a time, a squirrel and a dog were friends and always went about together. One day, they decided to get a sack and take themselves off to a grove to steal oranges. They left the house, and on arrival at the grove the squirrel said to the dog: 'Since I am small and supple, I will climb the tree. You stay down below to collect the fruit and put it in the sack.'

The squirrel climbed the tree and began to pick oranges. While picking them, the dog gathered them. Once the sack was half full, the squirrel directed an orange at the dog's head and hit it. The dog cried out, and the owners of the orange grove heard the dog and came running, shouting 'thieves! thieves!' as they came.

The squirrel quickly descended and asked the dog 'is it wise to cry out when one is in the middle of stealing something?' The squirrel wedged the dog's head between two trees and fled. The owners of the grove arrived and seized the dog, beating it to death.

The moral of this story is that theft is not a good thing.'

### Text 2: The story of Njh̄́́-mbuŋɔ́ (ATAK)

A true story recounted by Gaston Nguimeya, recorded by Nancy Haynes and Gretchen Harro, and transcribed by Pascal-Blaise Kemda.

Á le mmɔ́' ɨ́gu', mpfho njh̄́́-mbuŋɔ́, á wié mɔ́ menzwí yi, á gɔ́ méɨ́ tɛ́njú'aka. Mɔ́ menzwí a le tshwí́ tɛ́ ɨ́kū́' á lé la ɨ́gyā́, efɔ́ ntseɨ́alā́' á kɛ́. Njh̄́́-mbuŋɔ́, á piɨ́ njé́s pá' efɔ́ ntseɨ́alā́' á tsya a yí, ɨ́kɛ́ é á n̄́zwí yi lā́, á zaɨ́ a yí zā́ɨ́, á zā́ɨ́ a yí zā́ɨ́, tɛ́ pu ntóɨ́ ewɔ́ɨ́kh̄́́ e pí tɛ́ pu mm̄́, mbú́ n̄́cú'té́ á wɔ́p ewɔ́nzwí epí, mbú́ ndɔ́k efh̄́́ á fh̄́́, mbú́ ɨ́kā́́ álā́', n̄́cɛ́ é á mɛ́ndhú́. ɨ́gɔ́́ á mɛ́ndhú́, tɛ́ njé́s áju' nzéɨ́ n̄́nā́ɨ́ éwú. Efɔ́ a lejú́́ épū́p, á tsya a yí, á gɛ́ akézo mé ták á njh̄́́-mbuŋɔ́, tɛ́ pu ɨ́gɔ́́́ tɛ́ la'ne mé nzwɛ́t ezi ntsem-ntsem lé? Mé le shúɨ́́ á yí ɨ́gɛ́ mela'mū́ efɔ́ ntseɨ́alā́' a ke kɛ́ é á n̄́zwí yi. Efɔ́ a lejú́́ épū́p, mbɨ́ɨ́ ntó épū́p, ndé mé efɔ́ ntseɨ́alā́' ɨ́gɛ́: « á ka pɔ́ɨ́ ɨ́gɛ́, ó kɛ́ é á n̄́zwí njh̄́́-mbuŋɔ́. Yí mé tɛ́ sɨ́ ɨ́gɔ́́ á nda' menzwí éyi á n̄́tḗ á fh̄́́. Tsó' á n̄́zwí njh̄́́-mbuŋɔ́ ɨ́gyá́ á mbū́ yí » . Efɔ́ ntseɨ́alā́' a le jú' anu zi efɔ́ á le épū́p, n̄́zɨ́ɨ́ n̄́tsó' á n̄́zwí njh̄́́-mbuŋɔ́. Njh̄́́-mbuŋɔ́, á zeɨ́ efh̄́́ á mɛ́ndhú́ yi a ke pú́ ɨ́kɔ́́ n̄́tsé éwú lā́. Mbú́ zéɨ́ é́sɛ́t n̄́nɛ́t, n̄́zɨ́ɨ́ ɨ́gɔ́́́ mé mba ezi n̄́nā́ɨ́ éwú. Mé lɔ́k ndé ɨ́gɛ́: mé gɔ́ tɛ́ nták épū́p metsemé.

'The village elder njh̄́́-mbuŋɔ́ had taken a young fiancée, daughter of tɛ́njú'aka. The girl had now reached the marrying age when chief ntseɨ́alā́' took her illegally. In protest, njh̄́́-mbuŋɔ́ exiled himself and his large family to Bamendou.

<sup>5</sup>Note that this sentence is ungrammatical in the original French.

On learning of the departure of njhæ-mbuŋɔ, the paramount chief of Bafou was deeply saddened and asked himself what could have happened to njhæ-mbuŋɔ so that he left with all his family without saying goodbye. The paramount chief was informed that ntseŋalā' had taken the wife of njhæ-mbuŋɔ. On learning this, the paramount chief sent an emissary to ntseŋalā' to tell him that it was not acceptable to take the wife of njhæ-mbuŋɔ. 'She is not the only girl in Bafou. Return her to njhæ-mbuŋɔ.'

The chief ntseŋalā' obeyed the paramount chief and returned her. Njhæ-mbuŋɔ and his family returned to their ancestral property and re-occupied their compound.

The moral of this story is that you should know who it is that you are attacking.'

### Text 3: Sorcery (LEKAN)

This is a traditional story recounted by Maman Alice, recorded by Nancy Haynes and Gretchen Harro, and transcribed by Pascal-Blaise Kemda.

Mmɔ' menzwí a lelá' ígɔ́ éwú, mbɔ́ Ndém a lelá' ígyā́ á yí á ta' mɔ́. A le kɔ'té á mɔ́ mét, tɛ́ á wɛ́ ŋgu' tsɔp táa, megēm bíá, tɛ́ mbɔ́ á ku' a yí á lépu' ɛ́ athú á jka'te mmá yi. Megan lekaŋ é zwhi yí. Menzwí wɔ́, a lelá' ndá á méŋ yi mét ígu' lekua, té́ nná' ndēkné. Menzwí mét, a lelá' nnāŋ tɛ́ eshū́' amɔ́' álí'í, ígū́' á ésaá amɔ́' álā́'. A le kó éwū, ígīŋ á méŋ yi wɔ́, á gɔ́ ewú ntáná ésaá. A le jɔ́ pūp pá'ā, ndé eswāŋ nnēt, ígū́' álā́' pɔp, éfā́' á jgaŋ afho. Pó le gū́'ɔ́, tɛ́ ŋgḗ éshū́'í, mɔ́ wɔ́ á zɛ́ é á esaá. Jgaŋ afho wɔ́, á ŋɛtɛ́ tɛ́ ŋgɔ́ á yí, ndɔk ŋgū́' mé mba zi. Ndá' mbét á yí, tɛ́ á la' ntḗ, pɔ́ mmá yi, é gū́' mé mba zɔp. Tɛ́ mbú éfhó púp jkó áne eyá'a, mɔ́ mét a gɔ́ nzét á ígɔ́ŋ. Mé piŋ nzókó ánu ejú' mét, ndé ígɛ́: Séh ntḗ a tɛ́ pɔŋ.

'A woman lived with her only son. She raised the boy until he reached the age of 25, the age when he could help her. But the sorcerers killed him. The woman mourned her child for four years.

One day, the woman went to market in a far away village. In this market she caught sight of her son as he tried to escape her attention. So she returned to her own village and returned with a well-known healer.

On their arrival, they saw that the market was inhabited by the living dead. The healer seized the son and lead him away to treat him. Afterwards, the woman took her son and they returned home together. The son is still alive today.

The moral of this story is that jealousy is a flaw.'

### Text 4: The Monkey and the Panther (ANKA)

This is a traditional story recounted by Simplice Djoumessi, recorded by Nancy Haynes and Gretchen Harro, and transcribed by Pascal-Blaise Kemda.

Á lelá' ígɔ́ amɔ́' afu', na nzwi á nɔŋ ŋgū́' mé mba', mɔ́pɛ́t mena. A le gū́' tɛ́ ŋkó amɔ́' ajɔ́', ígwhū́ á ntḗ átā. A le náŋ wū́ élí'í lékua té́ pɛ́t ázɔ́. Anká a le shū́' amɔ́' álí'í, nnɔŋ ntókó ígū́' mé ŋkú' né, njú' ŋiŋ gɔ́ á ntḗ átā ígā́'té, á sū́tɛ́ njú' á gɔ́ na nzwi. Á zɛ́tɛ́ a yí ígɛ́: « Esó, ó nté átā íghū́é akézo? » Á gɛ́: Meŋ nde whū́é áta, á ghū́ élí'í lékua. Anká á zɛŋ ndé mé yí ígɛ́: « Meŋ efoko wú, o pɛ́t agá kɛ? » Na nzwi á gɛ́, ŋgáŋ. Anká á tsakne a asáŋ zí, nnéŋ á ntḗ átā, na nzwi tɛŋne ewú ntó. A le tó ntshī́é á ndó ŋgia púp pá'ā, ndé mé anká ígɛ́: « Anká, tɛ́ ó jɔ́ á légɛ́, meŋ nnáŋ á átā élí'í lékua té́ pɛ́t ázɔ́ mbij njú' á wú, ésaákné e? ». Anká á gɛ́: Tɛ́ o lɔk ndé ígɛ́ akézo? Á gɛ́ meŋ ā́ pɛ́t á wú. Pó le zɛ́é ŋaká, kúna á éshū́' nzétté á wɔp, ŋgɛ́ pɔ́ ŋaka akézo lé? Anká á pɛ́tɛ́ eshúŋ yí tɛ́ mi. Á ju' anuú mét, á kú'ne a yí tɛ́ á gɛ́: Pɛ́ zéŋ nízi, meŋ njú' pá' á ntók pūp, té́ ndɔk nzhé ágá léza' ntsaŋ. A le lé pūp pá'ā, na nzwi á zɛŋ esó' ó jkó á átā. Kúna á le ŋkú'té á ápú áŋká, ígɛ́ shū́'ḗ, mɔ́kɔ́ gū́'ɔ́, ecú leyhi égɔ́ égɔ́? Pó yí é le esáákné na nzwi, á ntḗ átā ígū́'ɔ́, té́ anká á lɔk mvū́' ó á léwhū́. Mé piŋ nzókó ánu ejú' mét, ndé ígɛ́: ŋiŋ a la' ŋghū́ á wú ánu mbɔŋ, tɛ́ tsɔ́' mé anu tɛpɔŋ.

'Once upon a time, the panther went into the forest to eat some of the other animals. In the forest, it suddenly found itself in deep pit, made for trapping animals. There the panther remained for four days without food or drink.

Then one day, the monkey was passing by and heard cries coming from the pit. The monkey went to look and saw that it was the panther. The monkey asked the panther: 'Dear friend, what are you doing in this pit?', to which the panther replied: 'I have been stuck here for four days!' Then the monkey asked: 'If I get you out of the pit, will you harm me?' But the panther remained silent. The monkey let its tail hang down into the pit and the panther caught onto it and climbed out.

The moment the panther was out, it said: 'Of course you have helped me greatly. But do you think I can go without food for four days and then just let you leave?' The monkey replied: 'What do you mean by that?' The panther responded: 'I am going to eat you.'

As they were speaking, the pig came by and asked them 'What are you discussing?' The monkey recounted the story. The pig asked the panther to re-enact the story from the beginning so that it could witness the events first-hand. Curiously, the panther agreed and climbed into the pit once more. The moment the panther was in the pit, the pig took the hand of the monkey and they departed. This time, the panther remained in the pit and died.

The moral of the story is as follows: never repay a good deed with a bad deed.'

## Reading data

```

ID tone marking (1=full, 0=zero)
| | story is familiar (1=familiar)
| | | text is anka
| | | | text is atak
| | | | | text is lekan
| | | | | | category of subject from cluster analysis (4=best, 1=worst)
| | | | | | | self-confidence (3=very confident, 1=under confident)
| | | | | | | | amount of primer completed (1.0 = all)
| | | | | | | | | education (2=some ter, 1=some sec, 0=some prim)
| | | | | | | | | | age
| | | | | | | | | | | sex (0=female, 1=male)
| | | | | | | | | | | | reading time
| | | | | | | | | | | | | reading disfluencies
| | | | | | | | | | | | | | tone performance errors
| | | | | | | | | | | | | | | tone comprehension errors
14 0 1 0 0 0 4 2 1.0 2 25 1 50.8 6.4 0 0
14 0 0 0 1 0 4 2 1.0 2 25 1 52.9 14.5 0 4
14 1 0 0 0 1 4 2 1.0 2 25 1 47.2 8.5 0 1
14 1 0 1 0 0 4 2 1.0 2 25 1 45.5 9.3 1 0
21 0 0 0 0 1 2 1 0.5 1 20 0 68.2 19.3 1 0
21 0 0 1 0 0 2 1 0.5 1 20 0 71.0 16.6 2 4
21 1 1 0 0 0 2 1 0.5 1 20 0 76.5 15.0 1 0
21 1 0 0 1 0 2 1 0.5 1 20 0 102.6 15.4 0 3
23 0 1 0 0 1 3 1 0.75 2 59 1 54.0 13.6 0 0
23 0 1 1 0 0 3 1 0.75 2 59 1 51.7 11.7 0 0
23 1 1 0 0 0 3 1 0.75 2 59 1 66.8 15.5 1 0
23 1 1 0 1 0 3 1 0.75 2 59 1 74.9 21.2 0 1
24 0 1 0 0 0 2 3 1.0 2 53 1 56.1 18.7 2 0
24 0 0 0 1 0 2 3 1.0 2 53 1 74.9 19.8 0 0
24 1 1 0 0 1 2 3 1.0 2 53 1 73.9 22.2 1 0
24 1 1 1 0 0 2 3 1.0 2 53 1 62.1 18.3 1 0
25 0 1 0 0 1 2 1 0.6 0 51 0 85.8 16.5 0 0
25 0 1 1 0 0 2 1 0.6 0 51 0 81.0 16.9 1 1
25 1 1 0 0 0 2 1 0.6 0 51 0 71.7 15.0 1 0
25 1 1 0 1 0 2 1 0.6 0 51 0 81.5 22.0 1 2
38 0 1 0 0 0 4 2 1.0 2 32 1 41.2 5.9 0 0
38 0 0 0 1 0 4 2 1.0 2 32 1 56.4 14.5 2 0
38 1 1 0 0 1 4 2 1.0 2 32 1 59.1 13.6 1 0
38 1 1 1 0 0 4 2 1.0 2 32 1 56.9 12.8 2 0
41 0 1 0 0 1 1 2 0.0 0 43 0 50.0 12.5 2 2
41 0 0 1 0 0 1 2 0.0 0 43 0 73.4 17.2 1 4
41 1 0 0 0 0 1 2 0.0 0 43 0 87.7 20.9 2 0
41 1 0 0 1 0 1 2 0.0 0 43 0 92.5 26.9 3 0
46 0 1 0 0 0 1 3 1.0 0 25 1 30.5 23.0 3 2
46 0 0 0 1 0 1 3 1.0 0 25 1 72.7 31.3 4 0
46 1 0 0 0 1 1 3 1.0 0 25 1 73.3 29.5 5 0
46 1 0 1 0 0 1 3 1.0 0 25 1 62.1 28.6 11 0
47 0 0 0 0 1 3 3 1.0 2 30 1 51.7 10.8 0 3
47 0 1 1 0 0 3 3 1.0 2 30 1 47.6 11.0 3 1
47 1 0 0 0 0 3 3 1.0 2 30 1 66.8 18.2 2 2
47 1 0 0 1 0 3 3 1.0 2 30 1 63.9 22.9 0 3
48 0 1 0 0 0 4 3 1.0 1 32 1 51.3 10.8 0 0
48 0 1 0 1 0 4 3 1.0 1 32 1 40.5 3.1 0 1
48 1 1 0 0 1 4 3 1.0 1 32 1 42.6 9.6 1 0
48 1 1 1 0 0 4 3 1.0 1 32 1 37.9 11.0 2 0
50 0 1 0 0 0 1 1 1.0 0 23 0 49.2 8.0 2 0
50 0 0 0 1 0 1 1 1.0 0 23 0 98.2 25.1 5 0
50 1 1 0 0 1 1 1 1.0 0 23 0 51.1 6.8 0 0
50 1 1 1 0 0 1 1 1.0 0 23 0 63.4 17.6 1 0

```