Neither deep nor shallow: a classroom experiment testing the orthographic depth of tone marking in Kabiye (Togo)

David Roberts, Stephen L Walter & Keith Snider

This pre-publication manuscript corresponds to the final, definitive version which has been published in *Language and Speech* 59:1.113–138 (2016) by SAGE Publications Ltd, All rights reserved.

Abstract

The experiment reported here tests the *Lexical Orthography Hypothesis*, that is, the notion that the output of the lexical phonology is the most promising phonological depth for an exhaustive representation of tone by means of diacritics in the orthography of a tone language. We conducted a controlled classroom experiment with 97 secondary school pupils learning written Kabiye, a Gur language of northern Togo. After testing their baseline skills in writing the standard orthography, the pupils participated in an eleven-hour transition course spread over three weeks in four parallel groups: DEEP (an experimental orthography representing the input of the lexical phonology), LEXICAL (representing the output of the lexical phonology), PHONEMIC (representing a level between the output of the lexical phonology and the output of the post-lexical phonology) and a control group. On the final day of the experiment, we tested their acquired skills in a dictation exercise. The results show that the LEXICAL group outperforms the other groups in three of the error types associated with adding diacritics, although they performed less well on one of the error types associated with writing long vowels. This initial evidence supporting the *Lexical Orthography Hypothesis* needs confirmation with reading and writing experiments on a variety of other tone languages.

Keywords

Tone, orthographic depth, lexical phonology, African languages, Kabiye, quantitative experiment

1. Introduction

Throughout the developing world, there are thousands of languages that do not yet have a written form. In many of these, tone plays a contrastive role in the lexicon and/or the grammar. The *World Atlas of Language Structures* plots not only a dense concentration of tone languages in sub-Saharan Africa and S.E. Asia, but also a broad scatter worldwide, including North, Central and South America, the Indian sub-continent, China, Papua New Guinea and Oceania (Maddieson 2013). By some estimates as many as 60-70% of the world's languages are tonal (Yip 2002, 1). Moreover, their geographical distribution loosely coincides with the very areas of the world in which mother-tongue illiteracy is most endemic. So orthography developers in these locations are frequently faced with the challenge of how to represent tone in writing.

In the course of the 20th century, numerous strategies were developed for marking tone in newly developed orthographies, including word-final silent letters in Romanizations of Asian languages (e.g. Heimbach 1969) word-initial and -final punctuation in Ivory Coast (e.g. Bolli 1978, 1991), and superscript numbers in Mexico (e.g. Rupp & Rupp 1996). But the classic, widespread convention for marking tone orthographically, particularly in sub-Saharan Africa, is the use of diacritics, and that method is the focus of this study.

The choice of symbolization entails an equally important but perhaps less obvious choice about which orthographic depth is most suitable: shallow or deep? The relative depth of different orthographies has been a key concept in the linguistics of writing ever since it was first introduced by Katz and Feldman (1981, 1983) and it remains the focus of ongoing research (e.g. Benuck & Peverly 2004; Caravolas *et al.* 2013; Snider 2014; Ziegler *et al.* 2010). Some orthographies, such as Spanish, are relatively shallow, mostly maintaining a one to one phonographic correspondence; others, such as French, are much deeper, the link between speech and spelling being less obvious.

Numerous researchers have debated what the most promising depth for a tone orthography should be (Bernard *et al.* 1995, 2002; Bird 1999a, 1999b; Koffi 1994; Kutsch Lojenga 2008, 2011; Mfonyam 1989, 1990) but as yet there is no consensus. This article seeks to shed new light on the subject by evoking the theory of lexical phonology. One of the most significant and lasting contributions this theory has made to our understanding of phonology is that there is a fundamental distinction between lexical and post-lexical linguistic processes and that native speakers are more psychologically aware of the former than they are of the latter. Although the theory itself has largely been superseded, some more recent theories, such as Stratal Optimality Theory (Bermúdez-Otero in preparation; Kiparsky 1998, 2000), continue to formally recognize the fundamental lexical/post-lexical distinction, so its applicability to orthography development remains intact. The current study reports on a classroom experiment that tested the *Lexical Orthography Hypothesis*, i.e. the notion that the output of the lexical phonology is the most promising depth for

an exhaustive representation of tone by means of diacritics. The experiment focuses on the orthography of Kabiye (ISO 639-3 kbp), an Eastern Gurunsi Gur language spoken in northern Togo.

We begin with some definitions. We will use the term "exhaustive" in this paper to describe any tone orthography that marks one less tone than the number of contrastive level tones in the language, that is, one tone in a language with two level tones and two tones in a language with three level tones. We consider such a representation to be exhaustive, or maximal, because no contemporary tone orthography literature advocates the writing of all tones everywhere. According to this definition then, all three experimental orthographies presented in this paper are exhaustive, because in each case H tone is marked by the presence of a diacritic and L tone by its absence. And from a methodological point of view, having all the experimental orthographies marked exhaustively kept the outcomes more comparable.

The term "downstep" also needs defining. In this paper, we use the term to describe a drop in tonal register. In Kabiye, as in many African languages, any H tone following a L tone is pronounced at a lower register than the preceding H, and this pattern reiterates until the end of the phonological phrase. Following Stewart's (1965, 1983, 1993) terminology, there are two manifestations of downstep: automatic (where the lowering of register is triggered by a pronounced L tone) and non-automatic (where the lowering of register occurs after a floating L tone, i.e. one that is not pronounced, but is present in the underlying structure). While some researchers, especially in the past, have used the terms downstep and downdrift to refer to these processes,

we prefer Stewart's terminology, which has gained wide acceptance, because it draws attention to a single process with two possible causes (for further terminological discussion, see Connell 2001).

The article is organized as follows. Section 2 provides a brief overview of the theory of lexical phonology and its implications for orthography development. Section 3 presents the Kabiye orthography: first some information about the sociolinguistic background, then a summary of the phonographic correspondences. Section 4 provides a sketch of the Kabiye tone system, and describes in detail the two tonal processes that are the focus of the experiment: lexical L tone spreading, and post-lexical HLH plateauing. Section 5 reports the experiment itself: the preparatory phase, the pre-test, the intervention, and the post-test. Section 6 presents the test results, treating the pre-test, an analysis of group equivalence, and each of the six error types in the post-test. Section 7 offers a summary and interpretation of the results, followed by some discussion of the implications of the experimental results for the development of tone orthographies for previously unwritten tone languages. Section 8 offers some concluding remarks.

H tone is written with an acute accent and L tone is written with absence of an accent in all three experimental orthographies and in phonetic transcriptions. In the latter, non-automatic downstep is written according to the IPA convention with a downward-pointing superscript arrow [ꜜ] placed before the affected tone bearing unit (TBU). In the experimental PHONEMIC orthography, this is replaced with an apostrophe <’>.

2. The Lexical Orthography Hypothesis

To understand the *Lexical Orthography Hypothesis*, a brief overview of lexical phonology and its implications for orthography development is necessary.

It is generally accepted that phonetic transcriptions of surface allophonic variations are not suitable for orthographic representation because such transcriptions represent sounds of which native speakers are almost totally unaware. For example, if the English orthography were to write both the aspirated and unaspirated allophones of /p/ (e.g., <spɪt> 'spit' vs. <pʰɪt> 'pit'), it would unnecessarily overdifferentiate, since native speakers are more or less unaware that the two variants of /p/ are actually different. The time-honored way of avoiding this extreme has been to employ *phonemic* representations, which avoid representing low-level allophonic differences like those in 'spit' and 'pit'. But while orthographies based on phonemic representations have often enjoyed relatively good success, consistently writing phonemic representations is also not without controversy.[1]

For example, native speakers are very aware of the outputs that result from lexical morphophonemic processes, and they have no problem

[1] Over the years, scholars have debated the relative merits of representing morphophonemic/underlying forms (cf. Chomsky and Halle 1968; Newman 1968; Venezky 1970) vs. phonemic forms (cf. Sapir 1933, 1949; Pike 1947; Nida 1954; Gudschinsky 1958, 1970, 1973). See also Snider 2014 for a discussion of this topic.

representing the phonemic outputs of those processes orthographically. Take, for example, the alternations of the English negative prefix /iN-/. While phonemic representations like <im-probable>, <ir-reverent>, and <il-logical> are not problematic for the orthography, morphophonemic representations like <in-probable>, <in-reverent>, and <in-logical> no doubt would be because native speakers are very aware that they pronounce these allomorphs differently. In fact, they are so aware of the differences that many do not even realize they are the same prefix.

On the other hand, native speakers are much less aware of phonemic outputs that result from postlexical morphophonemic processes, and they have no problem representing the outputs of those processes morphophonemically. Take, for example, the alternations between [s] and [z] of the English plural suffix in words like *cat*-*s* and *dog*-*z*. Even though /s/ and /z/ are phonemes in English (cf. *sip* vs. *zip*), native speakers are not very aware that they pronounce these phonemes differently in words like *cat*-*s* and *dog*-*z*, and they would probably not be very favorable to writing them differently in the orthography. Again, see Snider (2014, 36-40) for a discussion of problems associated with consistently representing the morphemic level orthographically.

The theory of lexical phonology (Goldsmith 1976/1990; Kiparsky 1982a, 1982b, 1985; Mohanan 1982, 1986), together with its more recent evolution into Stratal Optimality Theory (Bermúdez-Otero in preparation; Kiparsky 1998, 2000), makes a formal distinction between lexical and post-lexical phonological processes, with the former applying prior to the latter, in serial fashion. Following the lexical phonology model, Figure 1 illustrates the interaction

between morphosyntax and phonology, on the one hand, and the serial nature of the Lexical and Postlexical phonologies, on the other hand. We invert the classic diagram since, in terms of the linguistics of writing, it is metaphorically more intuitive for surface forms to appear at the top and deep forms at the bottom.

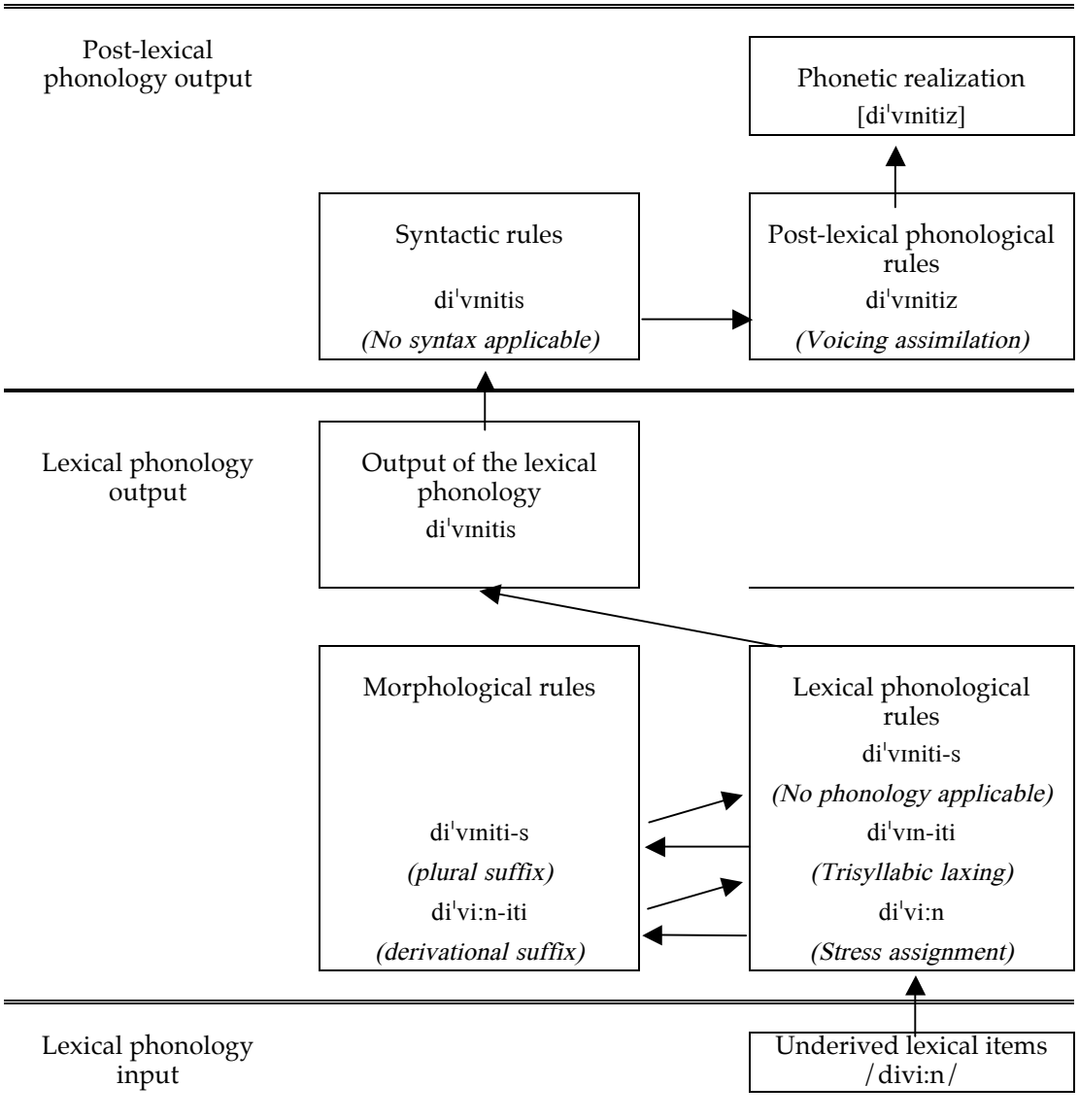

Figure 1. Lexical phonology model

Starting from the bottom of Figure 1, underived lexical entries are first subject to Lexical phonological processes. (The root of the example word 'divinities' is subject to stress assignment in the phonology side of the figure.) Subsequently, lexical entries may undergo any required Lexical morphological processes. (The example word undergoes derivational /–iti/ suffixation.) If lexical entries do undergo any morphological processes, the new, now morphologically complex form is again subject to any applicable Lexical phonological processes. (The example word now undergoes trisyllabic laxing, a process that shortens and laxes a long vowel when it is followed by two syllables, the first of which is unaccented.) This reciprocal interaction between morphology and phonology takes place repeatedly until all required lexical morphology has been affixed to the word and all applicable lexical phonological processes have applied to each new form. (The example word, morphologically, undergoes further suffixation of the plural suffix. Since this new form is not subject to any more phonology rules in the lexical component, nothing further happens to it phonologically in the lexical component.) The resultant output of all morphological and phonological processes in the lexical component is a "word." This (now derived) word is then subject to whatever phrasal syntax is appropriate to the communication intended by the speaker. Post-lexical phrases may be as simple as single-word utterances, or as complex as the syntax allows. (The example word is a single word pronounced in isolation, so no syntactic processes apply.) However simple or complex, once the phrase is assembled syntactically, it is subsequently subject to all applicable

post-lexical phonological processes. (The example word is now subject to the post-lexical Voicing Assimilation processes). For further discussion of how data is processed in the lexical phonology model, the interested reader is referred to Kenstowicz (1994, 196ff) and Kiparsky (1982a).

According to Mohanan (1982), native speakers are aware of lexical phonological processes and relatively unaware of post-lexical ones (Snider 2014, 30-36). If this is true, the output of the lexical phonology could be a promising level of phonological depth upon which to base an orthography.

Regarding post-lexical processes: a) they can result in non-contrastive sounds (e.g., aspirated consonants in English), b) native speakers can demonstrate a complete lack of awareness of phonological alternations that result from their application, and c) they can apply across word boundaries, with the result that some words may sound one way in one environment but another way in a different environment. Lexical processes, on the other hand: a) can have lexical or categorical exceptions (e.g., single words or whole classes of words that do not undergo the processes), b) can lack phonetic motivation, and c) can apply only word internally, including across morpheme boundaries.

Again, to the extent that native speakers are truly aware of phonological alternations that result from the application of lexical processes, and they are truly unaware of alternations that result from the application of post-lexical ones, the output of the lexical phonology promises to be an ideal beginning point for developing an orthography.

The theory of lexical phonology is usually applied to utterance-based linguistics, but we will apply it to the linguistics of writing by making the

following comparisons, from the shallowest (least abstract) to the deepest (most abstract) representations. In Table ***1***, the first column lists the terminology used in lexical phonology for each stage of the derivation, starting with the input at level 4 and ending with the pronunciation at level 1. The second column refers to the corresponding orthographic level, three of which (PHONEMIC, LEXICAL and DEEP) will be used throughout this paper to label the experimental tone orthographies. The third column gives a brief explanation of the phonographic relationship between symbol and sound at each level.

Table 1. Relationship between phonological and orthographic levels.

	Phonological level	**Orthographic level**	**Faithful to…**
1	Output of post-lexical phonology	SURFACE	…the pronunciation (phonetic transcription)
2	-	PHONEMIC	…the pronunciation minus application of any allophonic processes
3	Output of lexical phonology	LEXICAL	…the phonemic level minus application of any post-lexical processes
4	Input of lexical phonology	DEEP	…the lexical level minus application of any lexical processes. A morphographic representation.

Level (1) is the output of the post-lexical phonology, and corresponds to a phonetic transcription of the surface form. It includes non-contrastive phenomena such as automatic downstep and the falling pitch of L tones pre-pausally. This level is excluded from our study, because there is general agreement that an IPA transcription is not a valid candidate for an orthography. The bottom level (4) is the input of the lexical phonology. It corresponds to what is often called a DEEP orthography and is a morphographic representation, that is, the written representation of each morpheme remains unchanged

however it is pronounced. It is one of the representations that are included in the experiment.

Between these two extremes – surface and deep – are two other levels that are included in the experiment. They are often confused, because they are both relatively shallow, intermediate representations. First, there is the PHONEMIC level (2), which is close to the surface. Although it excludes non-contrastive tonal processes, it does include post-lexical processes whose outputs are phonemes in the language. This level does not figure in the theory of lexical phonology (which is why this level is blank in the first column of Table ***1***), but it is ripe for testing, because this is the level that the early structuralists (e.g. Gudschinsky 1973; Pike 1948; Swadesh 1934) advocated. And because of their enduring influence, many orthographies that are being developed today are still based on this level.

The other intermediate possibility is the LEXICAL level (3), which represents the output of the lexical phonology. While the structuralists systematically, and rightly, excluded from orthographic representation those post-lexical processes that resulted in non-contrastive sounds, they had no formal way of excluding post-lexical processes that did result in contrastive sounds, because this distinction was only identified later with the advent of lexical phonology. By excluding the output of *all* post-lexical processes in orthographic representation, the LEXICAL level not only excludes non-contrastive sounds but also those contrastive sounds that are below the awareness threshold of L1 speakers (i.e. that result from post-lexical processes). It is the usefulness of representing the LEXICAL level that is the focus of our

research. Examples of lexical and post-lexical processes in Kabiye will be given in Sections 4.2 and 4.3, respectively. With these distinctions in mind, we formulate the *Lexical Orthography Hypothesis* as follows:

Lexical Orthography Hypothesis

The lexical level (i.e., the output of the lexical phonology) offers the most promising level of phonological depth upon which to base a phonographic tone orthography that marks tone exhaustively.

This hypothesis is well known to some field linguists, but of course the only way to confirm or refute it is to formally test it. That is what motivated us to set up a quantitative classroom experiment, examining the exhaustive representation of tone by means of diacritics in the Kabiye orthography. We pitched three levels of orthographic representation against each other: DEEP, LEXICAL and PHONEMIC, with a view to determining which of these offered the most promising phonological level on which to base a diacritic tone orthography.

3. The Kabiye orthography

3.1. Sociolinguistic background

Kabiye is spoken by over one million people in Togo.[2] The Kabiye homeland consists of two mountain ranges to the north of the town of Kara, but

[2] Ethnologue (Lewis *et al.* 2015), citing Gblem-Poidi and Kantchoa (2012), estimates the Kabiye population to be 975,000. If this figure is accurate, and adding the annual 3.2% estimated growth rate for

the twentieth century saw considerable emigration to the centre and south of the country, so by now about two thirds of the population live outside the homeland.

The origins of the written form of Kabiye can be traced to the pioneering efforts of Rev. Antoine Brungard of the Société des Missions Africaines (1932). But it was not until the 1980s that the Comité de Langue Nationale Kabiyè (CLNK),[3] mandated by the Togolese government, devised the standard orthography that is still in use today. Standardization led to the development of a bilingual dictionary (CLNK & SIL-Togo 1999) and various government sponsored pedagogical materials (e.g. MAS 1984; MAS 1995). The catalogue of published literature now exceeds 200 titles. The national newspaper *Togo-Presse* contains a daily half page in Kabiye, and the new Kabiye Wikipedia – born in June 2014 – already has 150 articles in it.[4]

Various literacy initiatives exist in both the formal and non-formal sectors, including the option for pupils in state schools to take written Kabiye as an exam subject in grades nine and ten (Roberts 2011b). However, very few of these pupils ever actually encounter a Kabiye text, even in class. Togo is an overwhelmingly oral culture, and any literacy tends to be in French, the official

---

Togo (Alexandratos 1995), the Kabiye population will reach 1.07 million in the course of 2015.

[3] In 2013, the Comité de Langue Nationale Kabiyè changed its name to the Académie Kabiyè. We retain the former name in this article, since it was still being used at the time of our experiment.

[4] http://incubator.wikimedia.org/wiki/Wp/kbp/Main_Page. Accessed 20 April 2015.

language. In such a context, most readers can only be considered semi-literate at best, and this is certainly true of the pupils involved in our experiment.

Since there was early resistance among members of the CLNK and other literacy stakeholders to marking exhaustive tone with diacritics, the committee made only a single concession to tone marking: the disambiguation of a subject pronoun minimal pair (CLNK 1988, 9). However, as the years went by, the CLNK grew increasingly aware of the problems of zero tone marking, and have addressed this question in their meetings and their biannual journal (CLNK 1995, 4-5, 16-17). The principal author has participated in the debate over several years, contributing short, non-technical articles to the CLNK's journal, running a series of regional seminars with the aim of encouraging dialogue between orthography stakeholders, and leading a five-day extraordinary session of the CLNK during which various proposals were presented and discussed.

3.2.Phonographic correspondences

Kabiye is an SVO language with ten noun classes (Lébikaza 1999, 366-370) and vowel harmony (Lébikaza 1999, 65-69). It has 18 consonant phonemes /p, f, t, ʈ, s, t͡ʃ, k, k͡p, d, z, l, y, w, h, m, n, ɲ, ŋ/. The obstruent phonemes /p, f, t͡ʃ, k, k͡p/ are represented in the orthography either by the voiceless series <p, f, c, k, kp>, or by the voiced series <b, v, j, g, gb>, depending on various conditioning factors that are beyond the scope of this article but which have

been treated elsewhere (Lébikaza 1999, 135-140). As for the phonemes /j, ɲ/ they are represented by the graphemes <y, ñ> respectively.

There are nine basic vowel phonemes: /i, e, ɪ, ɛ, a, u, o, ʊ, ɔ/. The phonemes /ɪ, ʊ/ are represented respectively by the graphemes <ɪ, ʊ>. Long vowels are written as a sequence of two letters (<aa, ee> etc.), and extra-long vowels as a sequence of three (<aaa, eee> etc.). There is also a series of five long back unrounded vowels /ɯ̟ɯ̟, ɯ̟ɯ̟, ɤɤ, ʌʌ, ɑɑ/ that are contrastive with their back rounded and their front unrounded counterparts, but only occur at morpheme boundaries as the result of morphophonemic conditioning. They are written respectively as <iɣ, ɪɣ, eɣ, ɛɣ, aɣ>.

Finally, the hyphen <-> is used in the orthography to avoid homographs. It is placed between the possessive pronoun and the noun in the associative noun phrase (example 1), and between the verb and the object pronoun (example 2), to distinguish them from each other and from verb phrases (example 3):

	Prononciation	Standard orthography	
1	[ɛzáŋ]	<ɛ-saŋ>	*his showers*
	PP3/1-shower-4		
2	[ɛzáŋ]	<ɛsa-ŋ>	*he scratched you*

SP3s1-scratch-BP-OP2s

3 [ɛzáŋ] < ɛsaŋ > *he praises*

SP3/1-praise-IMP

4. The Kabiye tone system

4.1. Overview

Among Gur languages, Kabiye is relatively well described, with published research including not only two reference grammars (Delord 1976; Lébikaza 1999) but also dozens of other titles (see the thematic bibliography in Roberts 2013, 309-319). The tone system, too, is relatively well understood. All researchers agree that Kabiye has two contrastive level tones, high (H) and low (L), and automatic and non-automatic downstep (Lébikaza 1999, 183). The tone bearing unit (TBU) is the mora (Lébikaza 1999, 170, 267; Padayodi 2010, 249), which means that all vowels, and all preconsonantal and word final nasals are capable of bearing tone. Contour tones are not licensed on single TBUs. However, there are differing views about what constitute underlying forms. This paper takes the position that there are six tone patterns on noun roots – H, L, HL, LH, LHL and HLH (Roberts submitted) – and three on verb roots – H, L, HL – (Roberts 2002). We will discuss the implications of differing tone analyses for orthography development in Section 7.2.

Kabiye has tonal contrasts in both the grammar (examples 4 - 6) and the lexicon (7 - 8).

	Pronunciation	Standard orthography	
4	[ɛ-ɛ-tɔ-kí]	<ɛɛtɔkɪ>	*he is not eating*
	SP3sg_NC1-NEG-eat-IPF		
5	[ɛ-é-tɔ-kɪ]	<ɛɛtɔkɪ>	*when he will eat*
	SP3sg_NC1-DIS-eat-IPF		
6	[ɛ́-ɛ-tɔ-kí]	<ɛɛtɔkɪ>	*if he is not eating*
	SP3sg_NC1_CND-NEG-eat-IPF		
7	[kpó-ú]	<kpou>	*granary*
	granary-NC3		
8	[kpó-u]	<kpou>	*panther*
	panther-NC3		

This experiment focuses on two tonal processes: lexical L tone spreading (Section 4.2) and post-lexical HLH plateauing (Section 4.3). There are numerous others, but these were chosen because they are frequent in natural contexts, and because the distinction between their status as lexical and post-lexical processes is clear and unambiguous.

4.2.Lexical L tone spreading

The first tonal process that will be targeted for testing is lexical L tone spreading. In the Kabiye verb phrase, the L tone of a prefix spreads rightwards

onto a H verb root until it is blocked by a singly linked H tone.[5] In example *9*, the imperative (i.e. the base form of the verb) carries a H melody which associates to all three TBUs. In example *10*, the L tone of the subject pronoun causes the H tone on the first two tone bearing units of the verb root [wele-] to be pronounced as L, but has no effect on the third tone bearing unit [-sî]. In example *11*, the L tone of the negative prefix has exactly this same effect on its environment. Being a lexical process, the LEXICAL and PHONEMIC orthographies write L tone spreading in the same way, as it is pronounced, whereas the DEEP orthography remains faithful to the input of the lexical phonology, maintaining a fixed verb root image.

	Speech	DEEP orthography	LEXICAL and PHONEMIC orthographies	
9	[wélésî-∅]	<wélésî>	<wélésî>	*listen!*
	listen-IMP			
10	[e-welesî-na]	<ewélésîna>	<ewelesîna>	*he listened*
	SP3sg_NC1-listen_BP-COM			

[5] In autosegmental theory, tones and segments are represented on separate tiers, with association lines drawn between them to show how they interact. In the case of Kabiye, a "singly-linked H" refers to a single H tone pronounced on a single vowel or a nasal consonant.

11	[te-welesî-na]	< tewélésîna >	< tewelesîna >	*didn't listen*
	NEG-listen_BP-COM			

L tone spreading in Kabiye is considered to be a lexical process for three reasons. First, it applies internally, within the phonological word, never across a word boundary. Second, it is limited to a particular morphological context: not all L tones in the language spread to adjacent H tones, only those associated to verbal prefixes.[6] Third, classroom observations reveal that L1 speakers are often more aware of it than they are of post-lexical tonal processes – they hear it straightaway and learn how to write it more easily.

4.3.Post-lexical HLH plateauing

The second tonal process that will be targeted for testing in this experiment is post-lexical HLH plateauing. In Kabiye, a singly linked L between two H tones delinks, and the second H – with its downstepped register following a L – spreads left (Lébikaza 1999, 57-58). In example *12*, the melody of the word in isolation is [HL]. In example *13*, the melody is [HH]. When these

[6] Actually, a similar lexical process does occur in the associative noun phrase (Roberts 2003), but this is simply another example of a morphologically limited context, not a generalised pattern throughout the language. We do not refer to it here, because this construction was not targeted for testing in the experiment.

two words are juxtaposed in example *14*, this creates a HLH sequence, so the phrase is not pronounced [HLHH], but [H$^{\downarrow}$HHH]. Being a post-lexical process, the LEXICAL and DEEP orthographies write HLH plateauing in the same way, whereas the PHONEMIC orthography writes a level close to the output of the post-lexical phonology.

	Speech	DEEP and LEXICAL orthographies	PHONEMIC orthography	
12	[sɛ́-tʊ]	< sɛ́tʊ >	< sɛ́tʊ >	*thanks*
	thanks-NC9			
13	[fɛ́yí]	< fɛ́yí >	< fɛ́yí >	*there is not*
	there_is_no			
14	[sɛ́$^{\downarrow}$tʊ́ fɛ́yí]	< sɛ́tʊ fɛ́yí >	< sɛ́'tʊ́ fɛ́yí >	*don't mention it!*
	thanks-NC9 there_is_no			

An anonymous reviewer has suggested that the output of HLH plateauing is not phonemic but rather allophonic (with [HLH → H$^{\downarrow}$HH] and [HLLH → HLL$^{\downarrow}$H] as allophonic variants of underlying /HLH/), and that we are therefore not testing a phonemic orthography at this point. This, however, is not the case. From a structuralist point of view, the result of /HLH/ plateauing is [H$^{\downarrow}$HH], which contrasts phonemically with [HHH] and with [HLL]. Whether the

floating L responsible for non-automatic downstep is underlying or is generated lexically or postlexically is not a distinction structuralist theories are able to make, because they pre-date the theory of lexical phonology. Neither are they able to distinguish between singly and doubly linked L tones, because they pre-date the theory of Autosegmental Phonology.

HLH plateauing, and the non-automatic downstep that it generates, is considered to be a post-lexical process because it occurs in all environments,[7] within words and across word boundaries. It often occurs sequentially too, with the second H tone of one HLH sequence becoming the first H tone of the next HLH sequence. In the following four-word sentence (*15*), the HLH melody occurs five times, each time overlapping with the neighboring HLH pattern on either side.

	Speech	DEEP and LEXICAL orthographies	PHONEMIC orthography	
15	[ñá$^{\downarrow}$wéná $^{\downarrow}$égbé$^{\downarrow}$lé kú$^{\downarrow}$ɖúḿ $^{\downarrow}$ʈéké]	<ñáwená ɛgbélɛ kúɖʊḿ ɖeké>	<ñá'wéná 'égbé'lé kú'ɖúḿ 'ɖéké>	*as for you, you have only one maternal uncle*
	SP2sg_FOC be_BP-COM maternal.			

[7] In addition to this case of *post-lexical* non-automatic downstep, there are two cases of *lexical* non-automatic downstep in Kabiye, preceding two specific morphemes: the temporal clause marker [$^{\downarrow}$lé] and the conditional clause marker [$^{\downarrow}$yɔ́]. Again, we do not refer to them here because they were not targeted for testing in this experiment.

	uncle-NC1 one-NC1 only			

HLH plateauing is a generalized phenomenon throughout the language wherever the tonal context permits it, but classroom observations reveal that L1 speakers are often unaware of it – they do not necessarily hear it straight away, and learning to write it is a challenge. Another hallmark of HLH plateauing that identifies it as a post-lexical process is that non-fluent readers, whose reading is characterized by long hesitations between words, do not pronounce it across word boundaries, whereas fluent readers do.

4.4.Experimental predictions

Our experiment seeks to obtain empirical data providing evidence for or against the kinds of predictions made by the *Lexical Orthography Hypothesis*. For this purpose, we targeted the two tonal processes described in Sections 4.2 and 4.3 and developed three experimental orthographies to write them. Figure ***2*** summarizes the relationships between the two tonal processes and the three experimental orthographies, and our predictions about the relative ease of writing such forms.

Figure 2. Experimental predictions: Two tonal processes, three experimental orthographies.

Orthography	Lexical L tone spreading	Post-lexical HLH plateauing
PHONEMIC	Written as pronounced (easier)	Written as pronounced (harder)
LEXICAL		Written without post-lexical processes (easier)
DEEP	Written mophographically	

	(harder)	

Our predictions are that:

(a) those who learn the DEEP orthography will have more difficulty writing the correct tones on verb roots that undergo lexical L tone spreading, since it is difficult to abstract the underlying forms from the conjugated word.

(b) those who learn the PHONEMIC orthography will have difficulty writing post-lexical HLH plateauing, since it lies below the awareness threshold of mother-tongue speakers.

(c) The LEXICAL orthography is likely to be the only one of the three experimental orthographies that caters to both lexical and post-lexical processes adequately.

5. The experiment

5.1. Overview

The experiment took place on non-consecutive days from 26 January to 12 March 2011 at Lama-Kolidè secondary school in northern Togo.[8] Ninety-seven tenth-grade pupils participated in the experiment, all of whom had chosen written Kabiye as an optional exam subject. As one reviewer noted, written Kabiye may attract pupils who are naturally stronger in the language

[8] We would like to thank the headmaster, M. GARBA Tchandikou, for giving us access to his school.

arts, so the sample was not necessarily a representative cross-section of Kabiye pupils.

First, the pupils participated in two pre-intervention dictation tasks to establish their base-line skills in written Kabiye and French, after which we randomly assigned them to three parallel groups and a control group. Each group followed an eleven-hour course spread over three weeks. Finally, the pupils participated in a post-intervention Kabiye dictation test designed to assess their newly acquired skills. We also included a French dictation in the post-test, but it did not reveal any results relevant to the experiment.

5.2. Teacher training

We recruited eight mother-tongue teachers to teach the four courses in pairs. These were picked from a larger group of candidates for their demonstrated teaching ability and their facility with written Kabiye. The pedagogical materials were written, revised and introduced to the teachers by means of several rounds of pilot tests conducted with a small group of volunteers for several weeks in the run-up to the experiment. We closely observed the teachers' performance during this phase to ensure that they were well paired, and that the pairs would teach equally effectively across the three experimental groups. All the teaching was in Kabiye. The principal author was discretely present throughout the experiment, but only to ensure that it ran smoothly.

5.3.Preparatory phase

At the time of the experiment, the pupils had already completed approximately 40 hours of instruction in written Kabiye with a volunteer teacher who knows the standard orthography. We calculated this figure on the basis of 4 terms, 12 weeks per term, 1 hour a week, plus an estimated adjustment for the frequently interrupted school timetable. The pedagogy in such classes is characterized by rote learning and copying lessons directly from the blackboard into exercise books. There is a high pupil-teacher ratio (48 to 1). We considered this to be insufficient exposure to the written language to enable the pupils to participate in the experiment. So we supplemented their normal timetable with an additional two-hour lesson once a week for the five weeks preceding the experiment. During these lessons, the pupils benefitted from a much lower pupil-teacher ratio (11 to 1). Specially developed pedagogical materials gave them plenty of opportunity to practice writing and oral reading skills in the standard orthography. In addition, part of each lesson focused on applying the four basic conjugations (imperative, imperfective, unbound perfective and bound perfective) to a list of 20 verbs that would later be used in the dictation tasks.

During these preparatory weeks, each pupil completed a questionnaire in French enabling us to glean sociolinguistic data. In the frequent cases where the written responses proved to be insufficiently clear, a trained research assistant interviewed individual pupils in Kabiye, greatly increasing the accuracy of the data. This information was classified according to six independent variables (Table **2**).

Table 2. Independent variables gathered from the sociolinguistic questionnaire.

AGE	Pupil's age;
GENDER	Pupil's gender;
PARENT	Whether or not both of the pupil's parents were Kabiye;
DIASPORA	Whether or not the pupil attended primary school in a majority Kabiyephone region;
LANGUAGE	How many languages the pupil spoke;
READING	Estimated frequency of reading Kabiye, in addition to class and homework, measured in days per year.[9]

5.4.Pre-test

On the first day of the experiment, in a plenary session, we tested pupils' competence in the Kabiye standard orthography and the French orthography using dictation tasks composed of 20 short, semantically unrelated sentences. The inclusion of the French pre-test served primarily to identify any differences in pre-intervention skills that the pupils brought to the experimental setting. It also allowed us to control for any inequalities between the groups in terms of general writing ability. The variable of measurement was the number of individual spelling mistakes, quantified by comparing each pupil's written work letter by letter against two master texts (Kabiye and French) and counting

[9] Multiple choice answers to the question "How regularly do you read Kabiye in addition to class and homework?" were subsequently quantified (e.g. every day = 365; once a week = 52; once a month = 12; etc.) We asked a similar question with regard to frequency of writing, and learned that none of the pupils write in Kabiye in addition to classes and homework.

insertions, substitutions and omissions. We recorded their error rates by means of two independent variables (Table ***3***).

Table 3. Independent variables gathered from the pre-test.

KABIYE PRE-TEST	Raw error rate in the Kabiye standard orthography pre-intervention dictation test;
FRENCH PRE-TEST	Raw error rate in the French orthography pre-intervention dictation test.

5.5.Intervention

After the pre-test, the pupils were randomly divided into four parallel groups. Three of them learned the experimental orthographies, while a control group continued learning the standard orthography. These parameters were recorded by means of the grouping variable associated with the intervention, which we labeled GROUP (Table ***4***).

Table 4. Process variable GROUP.

DEEP	Those who learned the experimental DEEP orthography;
LEXICAL	Those who learned the experimental LEXICAL orthography;
PHONEMIC	Those who learned the experimental PHONEMIC orthography;
CONTROL	Those who continued to learn the standard orthography.

For an analysis and discussion of group equivalence see Section 6.2. Each group followed an eleven-hour training course spanning three weeks. The course consisted of eleven, one-hour lessons interspersed with practice dictation exercises.

5.6.Post-test

On the final day, we tested acquired skills in a Kabiye dictation task, with each group putting into practice whichever experimental orthography they had learned. The dictation contained 20 SVO sentences, similar in difficulty to those used in the pre-intervention task and the lessons, but different in content. The twenty sentences contained seven cases of lexical L tone spreading (see Section 4.2) and nine cases of post-lexical HLH plateauing (see Section 4.3). These sentences alternated with others that contained neither of these processes.

We recorded the results by means of six dependent variables. Table *5* explains each error type, describes how it was quantified and gives a concrete example of each type compared against a sentence written in the PHONEMIC experimental orthography: <Đájaa welesí'ná píɣá.> *Dadja listened to the child*

The incorrectly written TBUs are underlined. Other spelling errors, such as consonantal grapheme substitution, were not considered relevant to this experiment.

Table 5. Dependent variables – individual error types.

	Description	Measurement	Error example (PHONEMIC orthography)
ERROR1	Failing to write an acute accent on a TBU that should have one.	% of total number of TBUs that should be written with an accent.	<Đ<u>a</u>jaa welesí'ná píɣá.>

ERROR2	Writing an acute accent on a TBU that shouldn't have one.	% of total number of TBUs that should be written without an accent.	<Đáj**á**a welesí'ná píɣ́á.>
ERROR3	Writing a vowel where there shouldn't be one.	Raw error count.	<Đáj**a** welesí'ná píɣ́á.>
ERROR4	Failing to write a vowel where there should be one.	Raw error count.	<Đá**a**jaa welesí'ná píɣ́á.>
ERROR5	Failing to write lexical L tone spreading correctly.	Raw error count.	<Đájaa **wélés**í'ná píɣ́á.>
ERROR6	Failing to write non-automatic downstep in the context of postlexical HLH plateauing correctly.	Raw error count.	<Đájaa welesí**ná** píɣ́á.>

The counting of error rates related to vowel length (ERROR3 and ERROR4) in an experiment focusing on tone may seem of incidental interest. But in fact, it is particularly important to isolate these measures in a language like Kabiye that has contrastive vowel length, because if a short vowel is written as a long one, for example, <aa> instead of <a> (ERROR3), or vice versa, e.g. <a> instead of <aa> (ERROR4), there is no possibility that tone will be marked correctly. In other words, when length is concerned, accuracy in adding tone diacritics often depends on correctly spelling the vowels that bear them.

6. Analysis

6.1. Pre-test

Using the Minitab software, we ran ANOVAs for statistical variance to test whether any of the sociolinguistic independent variables (Tables ***2*** and ***3***) were predictive of pre-test performance in Kabiye or French.[10] The only one that was statistically significant was KABIYE PRE-TEST v. PARENT, $F(1,92) = 4.14$, $p = .045$.

Pupils write the Kabiye standard orthography with fewer errors if both parents are Kabiye. The high mean and low SD for the six pupils who had one non-Kabiye parent suggests consistent non-usage of Kabiye in these households. The much larger SD but lower error rate for the pupils who have two Kabiye parents reflect a high level of variation in use of Kabiye, but a lot of general familiarity with the language. None of the other independent variables – AGE, GENDER, DIASPORA, LANGUAGE and READING – were predictive of performance on the two pre-tests.

[10] One pupil was recorded as scoring nearly 100 percent more errors on the pre-test than any other pupil and 4 standard deviations above the mean. While we suspect a data entry error, we could not prove this and decided to retain this outlier in the analysis. This pupil was in the PHONEMIC group.

6.2. Group equivalence

Random assignment does not guarantee equivalence among groups especially when the number of participants in an experimental design is relatively small. Accordingly, further analysis was carried out to establish the level of group equivalence and to investigate the possible effects of any distributional anomaly that might have occurred. The results are captured in Tables **6** and **7**.

Table **6** defines the makeup of each group with respect to the distribution of the independent variables for which data was gathered. The four groups in the design are listed with demographic and performance characteristics by group shown by rows. P-values appear in the final column indicate whether the grouping characteristic under consideration was statistically significant.

	DEEP	LEXICAL	PHONEMIC	CONTROL	*p*
AGE	16.88	16.50	16.60	16.20	0.641
PARENT	91.7	95.5	95.7	92.0	0.913
DIASPORA	0.08	0.23	0.04	0.08	0.204
LANGUAGE	1.88	2.18	1.87	1.88	0.162
READING	17.00	7.82	2.09	14.72	0.119
KABIYE PRE-TEST	51.92	61.18	59.0	57.48	0.835
GENDER (F/M)	8 / 16	15 / 7	5 / 18	2 / 16	0.000
Mean KABIYE PRE-TEST score by GENDER (F/M)	63.6 / 46.1	69.2 / 44.0	48.4 / 61.9	111.0 / 52.8	

Table 6. Distributional characteristics of the four groups resulting from the random assignment of pupils.

The only characteristic that had a statistically non-random distribution with respect to GROUP was GENDER. The LEXICAL group was made up of a

disproportionally large number of female students compared to the other three groups. Clearly, this had the potential to perturb experimental findings if (a) a strong gender effect was found in the response variables or (b) there was evidence of interaction between GENDER and the interventions that defined each group. Indeed there is. The last row of Table **6** shows that girls made significantly more errors than did boys on the KABIYE PRE-TEST. Since the LEXICAL group has a disproportionate number of girls, it is possible that the performance of this group was affected, so further analysis was warranted.

This was carried out by (a) examining the impact of each sociolinguistic variable on pre-test performance, and (b) looking for evidence that the demographic variable in question impacted performance on the post-test in a manner independent of its impact on the pre-test. We included each variable in turn in a multivariate model (using GLM) with KABIYE PRE-TEST as a covariate. Table **7** shows the results of this investigation, with the key data in the final column. (For an explanation of each of the values, see Table **2**.

							p	Controlled *p*
AGE	14	15	16	17	18	19		
	54.4	43.2	54.0	65.2	75.9	53.3	0.210	0.871
PARENT	Y	N						
	55.4	85.2					0.045	0.701
DIASPORA	Y	N						
	62.9	56.6					0.598	0.069
LANGUAGE	1	2	3	4				
	63.8	52.5	79.8	78.0			0.120	0.842
READING	0	12	24	52	104			
	58.0	65.4	52.3	60.7	22.0		0.611	0.157
GENDER	F	M						
	67.0	52.7					0.067	0.640

Table 7. Investigation of the effects of independent variables on performance.

In Table 7, the columns to the left, which have specific labels on each line according to the values for each variable, give raw scores. For example, pupils who reported speaking only one language had a mean score on the pre-test of 63.8 while those reporting speaking two languages had a mean score of 52.5. The penultimate column contains *p*-values indicating whether variations in the variable in question can be considered statistically significant with respect to the pre-test. The final column shows *p*-values for each of the variables when included in a multivariate model with the pre-test as a covariate. A significant *p*-value in this case would indicate that the variable in focus impacted performance on the post-test independent of any impact on the pre-test.

The penultimate column reminds us that only one variable – PARENT – had a statistically significant impact on the pre-test. However, the *p*-value for PARENT in the multivariate model is large ($p = .701$), indicating that it did not independently impact performance on the post-test result.

In Table 6, it was noted that the LEXICAL group had a disproportionate number of female pupils. Given that females performed more poorly on the KABIYE PRE-TEST, we also tested for evidence of an interaction effect between experimental GROUP and GENDER using ANOVA. There was almost none ($p = 0.831$). This evidence plus the fact that the female pupils in the LEXICAL group performed at a level similar to other female pupils supports an interpretation that the gender profile of the four experimental groups did not have a significant impact on group response to the interventions introduced.

6.3. Post-test

The aim of the post-intervention dictation task was to compare proficiency in writing the three experimental tone orthographies. We did this by means of an analysis of each of the six individual error types against GROUP, the variable associated with the four groups. In each case, we quantified the results by counting individual deviations from these renderings in each pupil's written work when compared against three master texts, one for each experimental orthography.

6.3.1. Error type 1: Failing to mark an acute accent on a TBU that should have one.

An ANOVA analysis testing ERROR1 versus GROUP reveals that the LEXICAL and PHONEMIC groups function more or less as a single entity on this measure. The former outperforms the latter, scoring fewer errors, but the difference is not statistically significant. The DEEP group on the other hand scores far more errors than do the other two groups, $F(2,16)$, $p. < 0.001$.

6.3.2. Error type 2: Marking an acute accent on a TBU that shouldn't have one

An ANOVA analysis testing ERROR2 v. GROUP reveals, as before, that LEXICAL and PHONEMIC function more or less as a single entity. Again, the former outperforms the latter, scoring fewer errors, but the difference is not statistically significant. However, this time DEEP stands in marked contrast, scoring far fewer errors than do the other two groups on this same measure,

$F(2,66) = 14.18$, $p. < 0.001$, and in comparison with its own score on the ERROR1 measure.

But here we hit a measurement problem. If a pupil writes a TBU with no accent, is it a failure to write a H tone, a deliberate choice to mark L tone, or is it simply the result of fatigue, boredom, confusion or indecision? For this reason, it is misleading to measure ERROR2 in the same way as ERROR1.

So we examined the data from another angle, measuring how many sentences were written with at least one accent (Table *8*). We reasoned that a sentence written with no accents is most unlikely to represent a deliberate choice to write the whole sentence as L tone, since such sentences are almost non-existent in Kabiye. We also isolated the results of the first half of the dictation from the second half, to see if there was any decline in performance. The first line in Table *8* indicates, for each experimental orthography, the percentage of sentences written without any accents. The second and third lines contrast the average number of accentless sentences written in sentences 1-10 and sentences 11-20. The fourth line records the statistical probability of the results.

	Sentence #	DEEP	LEXICAL	PHONEMIC
% of sentences written without accents	1-20	36.3%	2.7%	14.5%
Average number of accentless sentences	1-10	2.125	0.318	1.182
Average number of accentless sentences	11-20	5.083	0.227	1.682
p-value		0.001	0.669	0.389

Table 8. Performance degradation measured by counting sentences written with no accents.

We observe that the DEEP group produced over a third of their sentences with no accents (in fact, only three pupils attempted to write at least one accent on every sentence). Moreover, there is a 139% increase in the number of sentences written with no accents between the first and second halves of the dictation. As for the PHONEMIC group, they fare better, but there are still many sentences written with no accents, and there is a 42% increase between the first and second halves of the dictation. In stark contrast, the LEXICAL group produced hardly any sentences with no accents (one pupil was responsible for almost half of them), and there is actually a decline in accentless sentences (-29%) between the first and second halves of the dictation, though this is not statistically significant. None of the other variables were predictive of performance on this measure.

Of course, viewing the data from this angle sheds no light at all on accuracy. Ironically, those in the DEEP group actually score fewer errors by giving up than the LEXICAL and PHONEMIC groups do by persisting to the end and continuing to make mistakes. But the results demonstrate that there is a marked tendency for the DEEP group to abandon their attempt to write the input of the lexical phonology, a trend that occurs much less in the PHONEMIC group's efforts to write a level near to the output of the post-lexical phonology, and is actually reversed in the LEXICAL group's experience of writing the output of the lexical phonology.

6.3.3. Error type 3: Writing a vowel where there shouldn't be one.

Running ANOVA to test ERROR3 v. GROUP does not deliver statistically significant results, indicating that there is no great variation between the four groups on the measure of writing short vowels as long.

6.3.4. Error type 4: Failing to write a vowel where there should be one.

Running ANOVA to test ERROR4 v. GROUP, reveals that the CONTROL group made far fewer errors than the other groups, $F(3,90) = 5.81$, $p. = 0.001$,[11] indicating that those writing the standard orthography write long vowels as short less frequently than the other groups. This is not particularly surprising since, as the control group, they were not laboring under the extra burden of marking accents, so they were free to concentrate on spelling the consonants and vowels correctly. What is more interesting is that the results of the three experimental groups are in inverse relationship to the ERROR2 degradation of performance rate discussed above, with the DEEP group performing best of the three experimental orthographies and the LEXICAL group faring worst. We interpret this to mean that the attention of the LEXICAL group as they learn to write the output of the lexical phonology is occupied in trying (albeit often failing) to mark the correct accents. This effect is less pronounced in the DEEP group writing the input of the lexical phonology. But this is not because their

[11] More surprisingly, GENDER also proves to be a strong predictor, with girls scoring many more errors than boys ($p = 0.000$). We have no explanation to offer for this tendency, which is not apparent in any of the other test results. See Section 6.2 for further discussion.

orthography is any easier. It is simply because so many of them have abandoned the task of adding diacritics altogether (see Section 6.3.2), and once they have, they can give their full attention to spelling the consonants and vowels correctly.

We should also note that both pre-tests are strong indicators of performance on ERROR4 (KABIYEPRE-TEST: correlation = .7; p = 0.000. FRENCHPRE-TEST: correlation = .226; p = 0.024), indicating that literacy skills in either language are a good basis for mastering the spelling of vowel length in Kabiye. This is the only case where we found a statistically significant correlation between the two pre-tests and one of the error types.[12]

[12] To measure the impact of the intervention on existing literacy skills, we also subtracted the scores of the two post-intervention tests from those of the two pre-intervention tests, and compared these results against GROUP. None of these results was statistically significant. However, a surprising result emerged from a comparison between the pre- and post-intervention French tests. Though different in content, these were similar in terms of difficulty and length, and we would not expect the mere three weeks of schooling that elapsed between the two tests to make a significant difference. So one would expect the pre-intervention test to be a strong predictor of the post-intervention test. This is true for two groups (PHONEMIC: correlation = .871, R-sq = .75.9; CONTROL: correlation = ..863, R-sq = .74.4), but not for the other two (LEXICAL: correlation = .545, R-sq = 29.7; DEEP: correlation = .539 R-sq = 29.1). Note the similar results within each pair, and the markedly dissimilar results between each pair. We consider this unusual enough to warrant reporting, though it is not central to the message of this paper and we have no explanation for it.

6.3.5. Error type 5: Failing to write lexical L tone spreading correctly

ERROR5 isolates one of the tonal processes under examination – lexical L tone spreading – and measures the error rates on the seven individually affected TBUs separately from the rest of the dictation text. We did this by grouping the experimental orthographies on the basis of how they write this tonal process.

The correct rendering of L tone spreading by the DEEP group is to mark an acute accent on a H root which is pronounced L, remaining faithful to the input of the lexical phonology. The correct rendering by the LEXICAL and PHONEMIC groups is to mark no accent, indicating the L tones as they are pronounced (see examples *9-11*, Section 4.2). It therefore makes sense to place the three orthographies into two groups, DEEP vs. LEXICAL+PHONEMIC, for this measurement. We have already observed a general difference between these two groupings, with many more errors being made by the DEEP group in error type 1 (Section 6.3.1), accompanied by a marked DEEP group degradation of performance rate in error type 2 (Section 6.3.2). However, when we focus on the individual TBUs affected by Lexical L tone spreading using the same binary groupings, they do not correlate or predict the outcome. The slight correlation between lexical and L tone spreading is almost entirely due to a correspondingly slight correlation between the pre-tests and the L tone spreading variable. This unclear picture is probably due to the measurement problem we have already noted (Section 6.3.2) with respect to marking L tone with absence of an accent.

6.3.6. Error type 6: Failing to write non-automatic downstep in the context of post-lexical HLH plateauing correctly.

ERROR6 isolates the post-lexical process targeted for the experiment. It measures error rates on the nine individual TBUs affected by non-automatic downstep in the context of HLH plateauing separately from the rest of the dictation text. As before, we did this by grouping the experimental orthographies on the basis of how they write this tonal process.

The correct rendering of non-automatic downstep in the context of post-lexical HLH plateauing by the DEEP and LEXICAL orthographies is the absence of an accent, representing the L tone of an underlying HLH sequence. The correct rendering by the PHONEMIC orthography is an acute accent over the TBU in question and an apostrophe preceding it (see examples *12-14*, Section 4.3).

So, as before, we placed the three orthographies into two groups. Examining the performance of DEEP+LEXICAL vs. PHONEMIC, the results are dramatic: the PHONEMIC group makes more than three times as many errors as the other grouping (p. = 0.000) and group membership accounts for 73.56% of the variance in the data. Adding the pre-tests as covariates made virtually no difference to the predictive relationship between these two variables, $F(1,67) = 186.36$, $p < 0.001$.

Until now, examining error types 1-5, the results of the LEXICAL group and the PHONEMIC group have shadowed each other: the difference between them has never been statistically significant. But suddenly in error type 6, there is a massive effect separating out these two groups. If a choice must be made

between the LEXICAL orthography (writing the output of the lexical phonology) and the PHONEMIC orthography (writing a level close to the output of the post-lexical phonology), there is no contest at least on this measure: the LEXICAL orthography is preferable. Writing non-automatic downstep is the weakest link in the PHONEMIC orthography.

7. Discussion

7.1. Summary and interpretation of results

Let us summarize the results of the analysis from the point of view of the pupils writing the LEXICAL orthography, since they were learning to write the output of the lexical phonology, which is the focus of the *Lexical Orthography Hypothesis*. Firstly, the LEXICAL group scored fewer errors writing an accent on a TBU that should have one than those writing the DEEP orthography (error type 1, Section 6.3.1), giving an indication that representing the output of the lexical phonology may be preferable to representing the input of the lexical phonology. The LEXICAL group also scored fewer errors writing post-lexical non-automatic downstep than those writing the PHONEMIC orthography (error type 6, Section 6.3.6), indicating that representing the output of the lexical phonology may be preferable to representing a level that contains some sounds that are present in the post-lexical phonology. The LEXICAL group also experienced less degradation of performance in the second half of the test than those writing the DEEP and PHONEMIC orthographies (error type 2, Section 6.3.2). But at the same time, they more frequently failed to write a vowel

where there should be one than the other groups (i.e. they often write long vowels as short), and this had the inevitable secondary effect of incorrect tone marking on these sequences (error type 4, Section 6.3.3). We interpret this to mean that the attention of the LEXICAL group as they learned to write the output of the lexical phonology was occupied in trying (albeit often failing) to mark the correct accents. This effect was less pronounced in the DEEP group writing the input of the lexical phonology. But this is not because their orthography was any easier. It is simply because so many of them abandoned the task of adding diacritics altogether, and once they did, they were able to give their full attention to spelling the consonants and vowels correctly.

We summarize four points supporting the *Lexical Orthography Hypothesis* in Table ***9***:

Table 9. Data supporting the Lexical Orthography Hypothesis.

Those writing the LEXICAL orthography -

1. scored fewer errors writing an accent on a TBU that should have one than those writing the DEEP orthography;
2. scored fewer errors writing post-lexical non-automatic downstep than those writing the PHONEMIC orthography;
3. experienced less degradation of performance on the second half of the task than those writing the DEEP and PHONEMIC orthographies;
4. were more absorbed with the task of writing accents correctly than those writing the DEEP and PHONEMIC orthographies (often causing them to write long vowels incorrectly).

In addition, we should add the observation – not mentioned until now – that, when all error types are collated in the Kabiye post-test, the LEXICAL group had three pupils who outperformed everyone else in the same group by a wide

margin. Checking back to see if this performance was replicated in the Kabiye and French pre-tests, we found that these pupils had performed well, but not outstandingly so. These three outliers in the Kabiye post-test could be considered as indicators of the potential of the LEXICAL orthography if pupils were given more opportunity for regular practice. If three above average pupils were able to master the LEXICAL orthography in an eleven-hour transition course spread over three weeks, it is likely that the other pupils in that same group would be able to replicate their performance if they had the benefit of more lessons and more practice time.[13] The PHONEMIC group had one such outstanding performer; the DEEP group had none.

7.2.The profile of the DEEP orthography

One of the major difficulties in developing an adequate DEEP orthography for Kabiye is that different researchers have come to different conclusions about what constitutes the underlying form of the verb, i.e. the input of the lexical phonology. This is a crucial issue, because a previous experiment has shown that verbs attract by far the most oral reading errors (Roberts 2010, 142). Lébikaza (1999, 215-221) identifies the underlying tone of the verb root by contrasting the tone in two contexts, the H tone adjectivizer [kí-] and the L tone locative [ɖɪ-], identifying eight underlying melodies: H,

[13] This was certainly the case when one of the authors taught his research assistants the LEXICAL tone orthography in preparation for a previous experiment with the double advantage of a longer training period and a low (2:1) pupil-teacher ratio (Roberts 2011a, 100-102).

HL, LH, LHL, LLH, L, LL, LLL. In discussion with the principal author, Lébikaza (personal communication, 2003) clarified that he considered it possible to group these into four basic melodies: H, L, HL and LH. Kassan (2000, 14) concurs with Lébikaza's published analysis. Padayodi (2010, 279-286) starts with the same prefix frames, but reduces the list to five: H, L, HL, LH, LHL. Roberts (2002, 36-39), on the other hand, considers the imperative to reveal the underlying form of the verb root, and acknowledges only three underlying melodies: H, L, HL. The DEEP orthography used in this experiment is based on the latter analysis.

Nouns present a similar scenario. Lébikaza (1999, 205-215), applying eight morphotonological rules, obtains seven underlying melodies: H, L, LL, LLL, HL, LH, LHL. Padayodi (2010, 259-274) identifies five underlying melodies: H, L, HL, LH and toneless. Roberts (submitted) on the other hand, obtains six underlying melodies: H, L, HL, LH, HLH, LHL. According to his analysis, the noun root melodies spread across the entire word and this is how the noun class suffix receives its tone. In other words, the input and output of the lexical phonology are identical for noun roots. Since the experiment followed Roberts' analysis, the DEEP and LEXICAL orthographies write nouns in the same way. We have already observed that pupils struggled to write the DEEP orthography; but it would be even more difficult to write DEEP tones in a representation where underlying noun root melodies are analyzed as being significantly different from surface forms.

Such conflicting analyses – by no means uncommon in tone research – are a reminder to us of one of the major hazards of developing DEEP

orthographies: identifying what the input of the lexical phonology is in the first place.

7.3.Depth and graphic density

All three tone orthographies developed for this experiment were 'exhaustive' or 'maximal' representations because they mark one less than the number of contrastive tones in the language. But we should also note that the choice of depth in a tone orthography may lead to representations of greater or lesser graphic density, even though all of them are exhaustive. Diacritic density is quantifiable by calculating the number of tone diacritics as a percentage of the number of tone bearing units in a natural text of about 400 words. (Bird (1999b, 89) uses 164 words, but our own recent research on several languages indicates that this is insufficient). In our experiment, the diacritic density of the DEEP orthography was 43%, that of the LEXICAL orthography was 48%, and that of the PHONEMIC orthography was 93%. The extremely high density of the latter is largely due to non-automatic downstep in the ubiquitous HLH sequences being written with only two acute accents in the DEEP and LEXICAL orthographies (e.g. **nákaú** *grasshopper*) but with three acute accents and an apostrophe in the PHONEMIC orthography (e.g. **ná'káú**). This disparity notwithstanding, the primary focus of this experiment was to explore orthographic depth not diacritic density, even though they are intertwined.

7.4.Accurately measuring 'absence of an accent'

Continuing on the topic of exhaustive tone marking, the binary strategy of marking H tone with an acute accent and L tone with absence of an accent created an unforeseen measurement challenge (Section 6.3.2). Every time a pupil wrote a TBU without an accent, it was unclear whether this represented failure to write a H tone or a deliberate choice to mark L tone. Two of the experimental groups (DEEP and PHONEMIC) also showed a marked decline in performance towards the end of the dictation. This concurs with Bird's (1999b, 101) observation in Dschang 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." Would the LEXICAL and PHONEMIC groups have succeeded better marking the L tones of lexical L tone spreading if they had been taught to mark it with a grave accent? Would the LEXICAL and DEEP groups have succeeded so well as they did in marking the L tone of post-lexical HLH plateauing if they had been taught to mark the affected TBU with a grave accent? These are questions that can only be asked with the benefit of hindsight.

What is clear is that only countable choices can be measured in a quantitative experiment. That is why we suggest that it may sometimes (depending on the nature of the research question, of course) be preferable to develop experimental orthographies that mark all tones everywhere. The purpose of such experiments, after all, is often to test a general orthographic *principle*, not a specific experimental orthography for its own sake.

The disadvantage of such an experimental strategy of course – as any literacy specialist will immediately point out – is that it is never advocated in the literature on developing orthographies for unwritten languages. It would be unfortunate if practitioners dismissed the lessons to be learned from an otherwise soundly designed experiment because the orthography used was deemed too far from field realities. There is a potential tension, then, between the quest to develop accurate and precise measurements, and the need for experiment designs to be ecological, capturing performance of an ability in way that is as true to real-life as possible.

7.5.Functional load and learner motivation

Languages differ enormously in the relative functional load carried by tone. Some Mande, Kru and Kwa languages in Côte d'Ivoire and some Omotic languages in Ethiopia are at the high end of the continuum (Wedekind 1985). These languages typically have four or five level tones, contour tones, and many monosyllabic words, all of which contribute to a high level of written ambiguity if tone is not marked. Previous research in Kabiye indicates that although the functional load of tone is not so low that it can be ignored, it is clearly not a language that warrants an exhaustive representation of tone.

Readers might legitimately ask, then, why we targeted Kabiye for investigation at all. The reason for the choice of language was purely pragmatic. Two of the authors had recently conducted a previous tone orthography experiment on the same language (Roberts & Walter 2012), so the professional contacts had already been nurtured in the language community, the

background linguistic research had already been done, and we could recycle many of the existing pedagogical materials.

But pragmatism apart, the relative functional load of tone in any given language is a crucial consideration for experimental research because it has implications for learner motivation. The tone orthography of a language with an extremely high functional load of tone may be more difficult to learn, but the learner has a correspondingly high motivation to overcome the obstacles because tone markings leads to comprehension. In a language with an extremely low functional load of tone, on the other hand, learner motivation is correspondingly low, because most of the diacritics are unnecessary and mastering the system pays no real dividends.

7.6.Methodological limitations

Our primary concern about the experimental methodology is that pupils' base-level skills in the standard orthography are weak. Since most of them have inadequate mastery of the segmental phonographic correspondence rules, they cannot be expected to add an extra graphic layer reliably. So the results show what performance is like when elementary pupils learn the simplest possible structures after minimal training, and should be understood with this in mind.

The experiment did not seek to assess oral reading skills. Yes, the results demonstrate that the LEXICAL orthography has certain advantages over a DEEP or a PHONEMIC representation, but they hold only for the writer, not the reader. And of course, what is an optimal orthography for one is not necessarily for the

other. Further research is needed to test the three levels of orthographic depth in oral reading experiments.

8. Conclusion

Fieldworkers engaged in developing orthographies for unwritten languages have many pressures on their time, and tone is often perceived as being of secondary importance, especially if the linguist in question is not an L1 speaker of a tone language. Many embark on tone orthography development with minimal training and support, unaware that the output of the lexical phonology offers a third option between the two extremes of PHONEMIC and DEEP forms. One of the aims of this article is to draw attention to the crucial need to disentangle lexical and post-lexical processes when deciding how tone should be spelled. Another aim is to underline the importance of distinguishing between the PHONEMIC and LEXICAL levels of orthographic representation, which none of the early literature was able to do because it predated the theory of lexical phonology. Since native speakers are less aware of post-lexical processes than lexical ones, post-lexical tonal processes should never be represented in a tone orthography.

But at a more fundamental level, this article carries no assumption that tone must necessarily be written exhaustively in all tone languages. On the contrary, we advocate that this strategy should only ever be used as a last resort, in languages where it has been proved that the functional load of tone is so high that it is unavoidable. Such proof comes not through autosegmental

theory – which is indispensable for analyzing the tone system itself – but from other less exploited angles such as analysis of written ambiguities in natural contexts and analysis of oral reading errors.

Who stands to benefit from the results of this experiment? Since we believe that the functional load of tone in Kabiye does not warrant an exhaustive representation of tone by means of diacritics, the results are almost certainly not going to be of much use to the CLNK as they debate how to mark tone beyond, perhaps, sending a clear signal to them about the complexities of orthographic depth inherent in exhaustive tone marking. In any case, the only members who advocate this strategy are one or two university trained linguists; most other members are intuitively opposed to it.

So is the experiment of purely scholarly interest to linguists versed in the intricacies of lexical phonology? No, because from the outset, it was conceived as responding to a global rather than a local concern. It looks beyond the borders of Togo and contributes to a wider debate about how tone should be written in orthographies throughout the world. With this in mind, we should bear in mind how vastly different tone systems can be from one another. It is clearly unwise to generalize about whether, how much, and by what means tone should be marked. This experiment provides clear evidence from one language that the output of the lexical phonology is the most promising level for an exhaustive diacritic tone orthography. This finding now needs to be further tested with multiple reading and writing experiments, characterized by a variety of imaginative aims and designs, in diverse languages from around the world.

9. Data tables[14]

9.1.Environmental variables

#	Group	Age	Gender	Parents	Diaspora	Language	Reading
1	Deep	18	F	0	0	2	0
2	Deep	15	F	1	0	1	0
13	Deep	16	M	1	0	2	0
15	Deep	18	F	1	0	2	0
20	Deep	16	M	1	0	2	52
21	Deep	16	M	1	0	1	12
22	Deep	16	M	1	0	2	0
23	Deep	17	M	1	0	1	0
26	Deep	17	M	1	0	1	0
27	Deep	20	M	1	0	2	0
28	Deep	20	M	1	0	2	0
30	Deep	16	F	1	0	1	0
34	Deep	14	M	1	0	2	18
37	Deep	22	F	1	0	2	0
42	Deep	15	F	1	0	2	24
46	Deep	16	M	1	1	2	24
59	Deep	18	M	1	0	3	0
69	Deep	16	M	1	1	2	104

[14] These data tables do not appear in the published version, but are available online on the Language & Speech website. Pupils 36, 40, 72 and 98 were absent for the final test.

73	DEEP	15	M	1	0	2	12
75	DEEP	17	F	1	0	2	0
77	DEEP	16	F	1	0	2	0
92	DEEP	16	M	1	0	2	52
93	DEEP	16	M	0	0	3	0
95	DEEP	19	M	1	0	2	24
5	CONTROL	17	M	1	0	2	12
8	CONTROL	15	M	1	0	2	52
9	CONTROL	17	M	1	0	2	0
14	CONTROL	16	M	1	0	2	0
16	CONTROL	16	F	1	0	2	0
19	CONTROL	17	M	1	0	2	24
29	CONTROL	16	M	1	0	2	0
35	CONTROL	15	M	1	0	1	0
38	CONTROL	17	M	1	0	2	0
45	CONTROL	14	M	1	1	1	12
48	CONTROL	17	M	0	0	3	0
49	CONTROL	16	F	1	0	2	0
52	CONTROL	17	M	1	0	2	24
53	CONTROL	18	M	1	0	2	0
55	CONTROL	16	M	1	0	1	52
58	CONTROL	15	M	1	0	1	0
64	CONTROL	16	M	1	0	2	12
67	CONTROL	15	M	1	0	1	24
82	CONTROL	15	M	1	0	1	0
85	CONTROL	16	M	1	0	2	0

86	CONTROL	16	M	1	0	2	0
88	CONTROL	16	M	1	0	2	104
91	CONTROL	16	M	1	0	3	0
94	CONTROL	17	M	1	0	2	52
97	CONTROL	19	M	0	1	3	0
6	PHONEMIC	18	M	1	0	2	0
7	PHONEMIC	18	M	1	0	2	0
10	PHONEMIC	17	M	1	0	2	0
11	PHONEMIC	17	M	1	0	2	0
25	PHONEMIC	17	M	1	0	1	0
31	PHONEMIC	16	M	1	0	2	0
32	PHONEMIC	17	F	1	0	2	0
33	PHONEMIC	15	M	1	0	2	0
39	PHONEMIC	15	M	1	0	2	24
43	PHONEMIC	22	M	1	0	1	0
56	PHONEMIC	18	M	1	0	2	0
62	PHONEMIC	16	M	0	1	3	0
63	PHONEMIC	17	F	1	0	2	0
70	PHONEMIC	16	M	1	0	1	0
71	PHONEMIC	16	F	1	0	2	0
79	PHONEMIC	14	M	1	0	2	0
80	PHONEMIC	15	F	1	0	1	0
81	PHONEMIC	16	M	1	0	2	0
83	PHONEMIC	17	M	1	0	2	12
87	PHONEMIC	15	F	1	0	2	0
89	PHONEMIC	17	M	1	0	2	0

90	PHONEMIC	16	M	1	0	1	0
96	PHONEMIC	19	M	1	0	3	12
3	LEXICAL	16	M	1	0	2	0
4	LEXICAL	17	F	1	0	2	0
12	LEXICAL	15	F	1	1	2	24
17	LEXICAL	16	F	1	0	3	52
18	LEXICAL	19	M	1	0	2	24
24	LEXICAL	19	F	1	0	2	0
41	LEXICAL	16	F	1	0	2	0
44	LEXICAL	15	F	1	0	2	0
47	LEXICAL	17	M	1	1	2	24
50	LEXICAL	16	F	1	0	2	0
51	LEXICAL	16	F	1	0	2	24
54	LEXICAL	16	M	1	0	2	24
57	LEXICAL	18	F	1	0	2	0
60	LEXICAL	16	F	1	0	2	0
61	LEXICAL	18	M	0	1	4	0
65	LEXICAL	18	F	1	0	2	0
66	LEXICAL	18	F	1	0	2	0
68	LEXICAL	14	F	1	1	2	0
74	LEXICAL	18	F	1	0	2	0
76	LEXICAL	15	F	1	1	3	0
78	LEXICAL	14	M	1	0	2	0
84	LEXICAL	16	M	1	0	2	0

9.2.Pre-test overall error rtes

#	GROUP	KABIYE PRE-TEST	FRENCH PRE-TEST
1	DEEP	82	18
2	DEEP	18	24
13	DEEP	18	31
15	DEEP	90	34
20	DEEP	39	39
21	DEEP	115	33
22	DEEP	17	24
23	DEEP	24	
26	DEEP	32	
27	DEEP	24	58
28	DEEP	44	22
30	DEEP	95	36
34	DEEP	21	21
37	DEEP	20	18
42	DEEP	43	17
46	DEEP	51	16
59	DEEP	81	16
69	DEEP	22	42
73	DEEP	28	21
75	DEEP	105	32
77	DEEP	56	24
92	DEEP	99	26
93	DEEP	73	

95	DEEP	49	25
5	CONTROL	31	29
8	CONTROL	49	21
9	CONTROL	72	18
14	CONTROL	25	24
16	CONTROL	119	23
19	CONTROL	45	
29	CONTROL	54	30
35	CONTROL	82	36
38	CONTROL	59	
45	CONTROL	73	34
48	CONTROL	114	19
49	CONTROL	103	17
52	CONTROL	81	
53	CONTROL	26	26
55	CONTROL	59	33
58	CONTROL	15	24
64	CONTROL	22	23
67	CONTROL	109	61
82	CONTROL	37	41
85	CONTROL	44	51
86	CONTROL	7	15
88	CONTROL	22	
91	CONTROL	31	14
94	CONTROL	65	
97	CONTROL	93	10

6	PHONEMIC	32	33
7	PHONEMIC	45	19
10	PHONEMIC	4	15
11	PHONEMIC	13	28
25	PHONEMIC	204	52
31	PHONEMIC	64	25
32	PHONEMIC	54	16
33	PHONEMIC	17	24
39	PHONEMIC	50	26
43	PHONEMIC	40	27
56	PHONEMIC	120	71
62	PHONEMIC	71	33
63	PHONEMIC	52	20
70	PHONEMIC	23	13
71	PHONEMIC	65	21
79	PHONEMIC	70	33
80	PHONEMIC	55	12
81	PHONEMIC	55	13
83	PHONEMIC	112	35
87	PHONEMIC	16	17
89	PHONEMIC	34	26
90	PHONEMIC	40	20
96	PHONEMIC	121	
3	LEXICAL	50	16
4	LEXICAL	64	
12	LEXICAL	14	15

17	LEXICAL	53	25
18	LEXICAL	28	18
24	LEXICAL	61	28
41	LEXICAL	47	22
44	LEXICAL	34	26
47	LEXICAL	74	13
50	LEXICAL	129	11
51	LEXICAL	75	
54	LEXICAL	9	11
57	LEXICAL	105	39
60	LEXICAL	51	21
61	LEXICAL	78	18
65	LEXICAL	51	38
66	LEXICAL	76	32
68	LEXICAL	72	14
74	LEXICAL	125	
76	LEXICAL	81	27
78	LEXICAL	36	16
84	LEXICAL	33	16

9.3.Post-test overall error rates

#	GROUP	KABIYE POST-TEST DIACRITICS	KABIYE POST-TEST CVS	FRENCH POST TEST
1	DEEP	77	56	13
2	DEEP	65	25	18
13	DEEP	66	28	4
15	DEEP	88	92	54

20	DEEP	80	43	34
21	DEEP	88	76	20
22	DEEP	60	13	24
23	DEEP	88	41	19
26	DEEP	63	38	40
27	DEEP	78	32	35
28	DEEP	62	21	10
30	DEEP	91	78	25
34	DEEP	60	27	18
37	DEEP	68	37	22
42	DEEP	72	47	26
46	DEEP	72	21	21
59	DEEP	76	26	19
69	DEEP	76	36	48
73	DEEP	80	51	16
75	DEEP	89	104	38
77	DEEP	85	66	
92	DEEP	88	79	28
93	DEEP	73	51	18
95	DEEP	73	79	30
5	CONTROL		23	28
8	CONTROL		50	15
9	CONTROL		49	17
14	CONTROL		20	18
16	CONTROL		106	22
19	CONTROL		30	17

29	CONTROL		41	14
35	CONTROL		74	28
38	CONTROL		53	34
45	CONTROL		79	26
48	CONTROL		115	14
49	CONTROL		97	20
52	CONTROL		47	13
53	CONTROL		27	18
55	CONTROL		36	24
58	CONTROL		18	18
64	CONTROL		28	7
67	CONTROL		77	66
82	CONTROL		35	50
85	CONTROL		35	37
86	CONTROL		5	3
88	CONTROL		16	3
91	CONTROL		44	11
94	CONTROL		51	62
97	CONTROL		70	19
6	PHONEMIC	100	53	37
7	PHONEMIC	118	57	15
10	PHONEMIC	18	22	8
11	PHONEMIC	94	31	24
25	PHONEMIC	104	76	70
31	PHONEMIC	88	51	30
32	PHONEMIC	113	70	16

33	PHONEMIC	92	41	13
39	PHONEMIC	93	54	21
43	PHONEMIC	100	57	22
56	PHONEMIC	103	73	81
62	PHONEMIC	96	82	34
63	PHONEMIC	82	63	25
70	PHONEMIC	86	33	8
71	PHONEMIC	84	53	16
79	PHONEMIC	99	86	66
80	PHONEMIC	95	53	16
81	PHONEMIC	102	38	28
83	PHONEMIC	108	71	50
87	PHONEMIC	89	26	7
89	PHONEMIC	95	56	14
90	PHONEMIC	115	38	30
96	PHONEMIC	108	84	41
3	LEXICAL	95	61	12
4	LEXICAL	75	72	31
12	LEXICAL	28	29	5
17	LEXICAL	95	70	17
18	LEXICAL	38	42	5
24	LEXICAL	95	98	
41	LEXICAL	79	54	23
44	LEXICAL	85	61	17
47	LEXICAL	55	72	16
50	LEXICAL	91	89	16

51	LEXICAL	97	59	19
54	LEXICAL	84	40	10
57	LEXICAL	91	87	28
60	LEXICAL	79	56	18
61	LEXICAL	94	91	33
65	LEXICAL	78	64	35
66	LEXICAL	93	76	27
68	LEXICAL	87	85	29
74	LEXICAL	103	103	39
76	LEXICAL	85	74	16
78	LEXICAL	88	43	11
84	LEXICAL	73	49	22

9.4.Post-test error types 1-4

#	GROUP	ERROR 1	ERROR 2	ERROR 2 (1-10)	ERROR 2 (11-20)	ERROR 3	ERROR 4
1	DEEP	84	22	3	9	0	12
2	DEEP	87	7	5	9	0	7
13	DEEP	90	6	4	8	0	4
15	DEEP	72	38	0	4	4	19
20	DEEP	67	33	0	0	3	9
21	DEEP	70	38	1	1	3	13
22	DEEP	76	8	3	5	1	2
23	DEEP	60	48	0	2	1	10
26	DEEP	81	9	4	10	1	8
27	DEEP	84	22	1	6	1	3
28	DEEP	85	5	4	7	0	3

30	DEEP	75	39	0	2	7	17
34	DEEP	78	8	2	8	6	3
37	DEEP	63	21	2	1	6	9
42	DEEP	88	13	2	10	1	16
46	DEEP	93	10	3	8	1	5
59	DEEP	67	31	0	7	0	5
69	DEEP	24	60	0	0	1	5
73	DEEP	46	45	0	0	6	8
75	DEEP	94	26	5	7	0	30
77	DEEP	73	33	2	6	3	10
92	DEEP	36	61	1	0	2	16
93	DEEP	99	7	8	10	1	8
95	DEEP	63	26	1	2	9	13
5	CONTROL					1	0
8	CONTROL					1	5
9	CONTROL					1	0
14	CONTROL					2	1
16	CONTROL					12	10
19	CONTROL					2	7
29	CONTROL					3	8
35	CONTROL					13	7
38	CONTROL					3	8
45	CONTROL					1	16
48	CONTROL					5	15
49	CONTROL					0	17
52	CONTROL					5	8

53	CONTROL					1	6
55	CONTROL					1	9
58	CONTROL					2	6
64	CONTROL					0	1
67	CONTROL					2	10
82	CONTROL					4	6
85	CONTROL					0	7
86	CONTROL					0	1
88	CONTROL					1	2
91	CONTROL					6	4
94	CONTROL					2	6
97	CONTROL					4	8
6	PHONEMIC	48	65			8	7
7	PHONEMIC	47	88	0	0	3	8
10	PHONEMIC	18	4	0	0	1	2
11	PHONEMIC	41	70	0	0	2	6
25	PHONEMIC	54	69	1	2	1	22
31	PHONEMIC	78	29	4	1	3	12
32	PHONEMIC	51	72	0	0	16	10
33	PHONEMIC	65	44	1	3	4	3
39	PHONEMIC	54	51	0	1	5	15
43	PHONEMIC	86	31	3	6	9	13
56	PHONEMIC	63	60	3	1	2	24
62	PHONEMIC	33	76	0	0	3	23
63	PHONEMIC	57	40	1	1	2	8
70	PHONEMIC	70	29	1	3	4	8

71	PHONEMIC	53	52	0	2	1	9
79	PHONEMIC	85	34	1	8	5	22
80	PHONEMIC	68	44	6	3	6	11
81	PHONEMIC	58	62	0	2	4	6
83	PHONEMIC	30	92	0	0	5	20
87	PHONEMIC	76	28	4	3	6	4
89	PHONEMIC	59	48	1	1	5	10
90	PHONEMIC	39	90	0	0	5	8
96	PHONEMIC	61	56	0	0	1	17
3	LEXICAL	38	71	0	0	0	13
4	LEXICAL	71	24	0	0	1	19
12	LEXICAL	15	18	0	0	2	0
17	LEXICAL	32	66	0	0	10	11
18	LEXICAL	14	29	0	0	3	12
24	LEXICAL	71	46	0	0	4	26
41	LEXICAL	42	51	2	0	2	9
44	LEXICAL	33	61	0	0	3	16
47	LEXICAL	19	43	0	0	0	13
50	LEXICAL	79	36	0	0	0	28
51	LEXICAL	74	46	2	3	5	10
54	LEXICAL	50	49	0	0	2	2
57	LEXICAL	78	35	1	0	2	20
60	LEXICAL	43	49	0	0	2	11
61	LEXICAL	60	53	0	0	5	16
65	LEXICAL	25	57	0	0	14	10
66	LEXICAL	82	33	1	2	5	17

68	LEXICAL	67	39	0	0	4	14
74	LEXICAL	33	83	0	0	0	26
76	LEXICAL	46	55	0	0	3	22
78	LEXICAL	43	60	0	0	0	12
84	LEXICAL	67	26	1	0	2	5

9.5. Post-test error types 5-6

LEXICAL: Grouping of the three experimental orthographies for lexical processes (DEEP is written one way, LEXICAL & PHONEMIC are written another).

POSTLEXICAL: Grouping of the three experimental orthographies for post-lexical processes (PHONEMIC is written one way, DEEP and LEXICAL another).

#	GROUP	LEXICAL	POSTLEXICAL	L tone spreading	HLH plateauing
1	DEEP	DEEP	DEEP&LEXICAL	5	1
2	DEEP	DEEP	DEEP&LEXICAL	4	1
13	DEEP	DEEP	DEEP&LEXICAL	8	1
15	DEEP	DEEP	DEEP&LEXICAL	6	2
20	DEEP	DEEP	DEEP&LEXICAL	9	4
21	DEEP	DEEP	DEEP&LEXICAL	7	3
22	DEEP	DEEP	DEEP&LEXICAL	4	1
23	DEEP	DEEP	DEEP&LEXICAL	5	3
26	DEEP	DEEP	DEEP&LEXICAL	7	0
27	DEEP	DEEP	DEEP&LEXICAL	5	1
28	DEEP	DEEP	DEEP&LEXICAL	6	1
30	DEEP	DEEP	DEEP&LEXICAL	3	1
34	DEEP	DEEP	DEEP&LEXICAL	4	1
37	DEEP	DEEP	DEEP&LEXICAL	4	1

42	DEEP	DEEP	DEEP&LEXICAL	6	0
46	DEEP	DEEP	DEEP&LEXICAL	7	1
59	DEEP	DEEP	DEEP&LEXICAL	5	1
69	DEEP	DEEP	DEEP&LEXICAL	0	4
73	DEEP	DEEP	DEEP&LEXICAL	5	5
75	DEEP	DEEP	DEEP&LEXICAL	7	1
77	DEEP	DEEP	DEEP&LEXICAL	6	3
92	DEEP	DEEP	DEEP&LEXICAL	1	4
93	DEEP	DEEP	DEEP&LEXICAL	8	1
95	DEEP	DEEP	DEEP&LEXICAL	1	2
6	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	8	8
7	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	7	6
10	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	0	4
11	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	6	7
25	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	9	6
31	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	4	7
32	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	9	9
33	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	4	6
39	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	3	7
43	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	0	9
56	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	6	7
62	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	7	7
63	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	7	8
70	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	5	7
71	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	8	7
79	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	2	7

80	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	3	7
81	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	5	6
83	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	7	6
87	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	2	8
89	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	8	8
90	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	9	8
96	PHONEMIC	LEXICAL&PHONEMIC	PHONEMIC	3	6
3	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	5	4
4	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	4	3
12	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	2	2
17	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	8	3
18	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	4	1
24	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	3	6
41	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	2	2
44	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	6	4
47	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	2	0
50	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	5	4
51	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	4	3
54	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	3	5
57	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	2	3
60	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	8	3
61	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	2	2
65	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	3	3
66	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	4	2
68	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	8	1
74	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	9	3

76	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	8	3
78	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	4	1
84	LEXICAL	LEXICAL&PHONEMIC	DEEP&LEXICAL	3	0

Abbreviations

BP	bound perfective
CLNK	Comité de Langue Nationale Kabiyè (now the Académie Kabiyè)
CND	conditional
COM	comitative
DIS	distant
H	high tone
IMP	imperative
IPF	imperfective
L	low tone
NC	noun class
NEG	negative
PP	possessive pronoun
sg	singular
SP	subject pronoun
SVO	subject, verb, object
TBU	tone bearing unit

Acknowledgements

We are deeply indebted to the principal author's research assistant, Emmanuel Pidassa, without whose efforts this research would never have been completed. We would like to thank Mike Cahill, Steve Parker and four anonymous reviewers for their stimulating and helpful comments on a draft of this article. We are indebted to the CLNK for the interest they have taken in this research since it began. The article also benefitted from the discussion following its presentation in English at the 41st Colloquium on African Languages and Linguistics (University of Leiden, Netherlands, 29-31 August 2011) and in French at the LLACAN-CNRS Journée Scientifique, (Villejuif, France, 18 November 2011).

Funding

This research was made possible thanks in part to a twelve month postdoctoral research fellowship awarded to the principal author by the City of Paris to work at the LLACAN-CNRS research laboratory.

References

Alexandratos, N. (Ed.) (1995). *World Agriculture Towards 2010: An FAO Study*. New York: Food and Agriculture Organization of the United Nations and John Wiley and Sons.
Benuck, M. B., & Peverly, S. T. (2004). The effect of orthographic depth on reliance upon semantic context for oral reading in English and Hebrew. *Journal of Research in Reading*, 27. 3, 281-299.
Bermúdez-Otero, R. (in preparation). *Stratal Optimality Theory*. Oxford: Oxford University Press.
Bernard, R. H., Mbeh, G. N., & Handwerker, W. P. (1995). The tone problem. In A. Traill, R. Vossen, & M. Biesele (Eds.), *The complete linguist: papers in memory of Patrick J. Dickens* (pp. 27-44). Cologne: Rüdiger Köppe.
Bernard, R. H., Mbeh, G. N., & Handwerker, W. P. (2002). Does marking tone make tone languages easier to read? *Human organisation*, 61. 4, 339-349.
Bird, S. (1999a). Strategies for representing tone in African writing systems. *Written Language and Literacy*, 2. 1, 1-44.
Bird, S. (1999b). When marking tone reduces fluency: an orthography experiment in Cameroon. *Language and Speech*, 42. 83-115.
Bolli, M. (1978). Writing tone with punctuation marks. *SIL Notes on Literacy*, 23. 16-18.
Bolli, M. (1991). Orthography difficulties to be overcome by Dan people literate in French. *SIL Notes on Literacy*, 65. 25-34.
Brungard, A. (1932). *Taakayã kɔɔnɔŋga*. Lomé: Mission Catholique, vicariat apostolique du Togo.
Caravolas, M., Lervåg, A., Defior, S., Málková, G. S., & Hulme, C. (2013). Different Patterns, but Equivalent Predictors, of Growth in Reading in Consistent and Inconsistent Orthographies. *Psychological Science*, 24. 8, 1398-1407.
CLNK (1988). *Ɛbɛ Laba ? (Qu'est-ce qui se passe ?) : la revue semestrielle du comité de langue nationale kabiyè, no. 5*. Lomé: CLNK.
CLNK (1995). *Procès verbal de la deuxième session ordinaire annuelle du 6 au 10 novembre 1995 du comité de langue nationale kabiyè*. Lomé: CLNK.
CLNK, & SIL-Togo (1999). *Tɔm kpou kabɩyɛ-fransɩɩ (Dictionnaire kabiyè-français)*. Lomé: Ministère de l'Éducation Nationale et de la Recherche, Comité de Langue Nationale Kabiyè et SIL-Togo.
Connell, B. (2001). *Downdrift, Downstep, and Declination*. Paper presented at the Typology of African Prosodic Systems Workshop. Bielefeld University, Germany.
Delord, J. (1976). *Le kabiyè*. Lomé: Institut national de la recherche scientifique.
Gblem-Poidi, H., & Kantchoa, L. (2012). *Les langues du Togo : Etat de la recherche et perspectives*. Paris: Harmattan.
Goldsmith, J. (1976/1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.
Gudschinsky, S. C. (1973). A Manual of Literacy for Preliterate Peoples In (pp. 1-180). Dallas: SIL Ukarumpa. Also available in SIL International Digital Resources LinguaLinks library 5.0 plus.
Heimbach, E. E. (1969). *White Hmong-English Dictionary*. New York: Cornell University, Ithaca.
Kassan, B. B. (2000). De l'influence du ton consécutif dans les formes de l'aoriste en kabiyè. *Cahiers voltaïques / Gur Papers*, 5. 13-22.

Katz, C., & Feldman, L. B. (1981). Linguistic Coding in Word Recognition: Comparisons between a Deep and a Shallow Orthography. In A.M. Lesgold, & C.A. Perfetti (Eds.), *Interactive Processes in Reading* (pp. 85–106). Hillsdale, NJ: Erlbaum.
Katz, C., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9. 157–166.
Kiparsky, P. (1982a). From cyclic to lexical phonology. In H. Van Der Hulst, & N. Smith (Eds.), *The structure of phonological representations* (pp. 131-175). Dordrecht: Foris.
Kiparsky, P. (1982b). Lexical phonology and morphology. In I.S. Yang (Ed.) *Linguistics in the Morning Calm* (pp. 3-91). Seoul: Hanshin.
Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology Yearbook*, 2. 83-138.
Kiparsky, P. (1998). Paradigm Effects and Opacity (manuscript). Stanford University
Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review*, 17. 351- 367.
Koffi, E. (1994). The representation of tones in the orthography. *SIL Notes on Literacy*, 20. 3, 51-9.
Kutsch Lojenga, C. (2008). *The orthography of tone in African languages*. Paper presented at the 5th symposium "The idea of writing: the unspoken parts of writing". University of Leiden, Netherlands.
Kutsch Lojenga, C. (2011). *Orthography and Tone: Tone system typology and its implications for orthography development* Paper presented at the Annual meeting of the Linguistics Society of America. Pittsburgh, Pennsylvania.
Lébikaza, K. z. K. (1999). *Grammaire kabiyè: une analyse systématique - phonologie, tonologie et morphosyntaxe*. Cologne: Rüdiger Köppe.
Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.) (2015). *Ethnologue: Languages of the World, 18th edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.
Maddieson, I. (2013). Tone. In M.S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology (Available online at http://wals.info/chapter/13, Accessed on 2014-11-01.).
MAS (1984). Nakaa nɛ Kpacaa : takayaɣ kɩkɛlaɣ. In *Naka et Kpatcha : syllabaire kabiyè, 2e livre, apprenant*. City: Commission régionale de langue kabiyè du ministère de la santé publique, des affaires sociales et de la condition féminine avec le concours de la SIL.
MAS (1995). Nakaa tɛ : Kɔɔnaɣ takayaɣ. In *Chez Naka : syllabaire kabiyè, 1e volume, apprenant*. City: Commission régionale de langue kabiyè et SIL.
Mfonyam, J. N. (1989). *Tone in Orthography: the Case of Bafut and Related Languages*: Université de Yaoundé, Cameroun. Thèse d'état.
Mfonyam, J. N. (1990). Tone analysis and tone orthography. *Journal of West African Languages*, 20. 2, 19-30.
Mohanan, K. P. (1982). *Lexical Phonology*. Cambridge, MA: MIT
Mohanan, K. P. (1986). *The theory of lexical phonology*. Dordrecht: Reidel.
Padayodi, C. c. M. (2010). *A revised phonology of Kabiye segments and tones*. Arlington, TX: University of Texas at Arlington. PhD dissertation.
Pike, K. (1948). *Tone Languages: a Technique for Determining the Number and the Type of Pitch Contrasts in a Language*. Ann Arbor: University of Michigan press.
Roberts, D. (2002). *Les classes tonales du verbe en kabiyè*: Université de la Sorbonne nouvelle, Paris III. Mémoire de maîtrise.
Roberts, D. (2003). Tone spreading in the Kabiye associative noun phrase. *Cahiers voltaïques / Gur Papers*, 6. 95-100.
Roberts, D. (2010). Hidden morpheme boundaries in Kabiye: a source of miscues in a toneless orthography. *Writing Systems Research*, 2. 2, 139–153.
Roberts, D. (2011a). Autosegmental and pedagogical considerations in preparation for a tone orthography experiment. *Journal of West African Languages*, 38. 2, 87-106.
Roberts, D. (2011b). The development of written Kabiye and its status as one of the "national" languages of Togo. In O. Orwenjo, & O. Ogone (Eds.), *Language and politics In Africa: Contemporary issues and critical perspectives* (pp. 468-494). Newcastle-on-Tyne: Cambridge Scholars Publishing.
Roberts, D. (2013). *Parlons kabiyè*. Paris: Harmattan.

Roberts, D. (submitted). Root melodies associating to whole words: a tone analysis of Kabiye nouns. *Journal of African Languages and Linguistics*.
Roberts, D., & Walter, S. L. (2012). Writing grammar rather than tone: an orthography experiment in Togo. *Written Language & Literacy*, 15. 2, 226-253.
Rupp, J. E., & Rupp, N. (1996). *Diccionario chinanteco de San Juan Lealao Oaxaca*. Tucson: SIL.
Snider, K. (2014). Orthography and phonological depth. In M. Cahill, & K. Rice (Eds.), *Developing Orthographies for Unwritten Languages* (pp. 27-48). Dallas: SIL International.
Stewart, J. M. (1965). *The typology of the Twi tone system*. Legon, Ghana: Institute of African Studies, University of Ghana.
Stewart, J. M. (1983). Downstep and floating tones in Adioukrou. *Journal of African Languages and Linguistics*, 5. 1, 57-78.
Stewart, J. M. (1993). Dschang and Ebrie as Akan-type total downstep languages. In H. Van der Hulst, & K. Snider (Eds.), *The Phonology of Tone – the representation of tonal register* (pp. 185– 244). Berlin: Mouton de Gruyter.
Swadesh, M. (1934). The phonemic principle. *Language*, 10. 118-129.
Wedekind, K. (1985). Thoughts when drawing a map of tone languages. *Afrikanistische Arbeitspapiere*, 1. 105-124.
Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
Ziegler, J. C., Bertrand, D., Tóth, D. n., Csépe, V. r., Reis, A., Faîsca, L. s., Saine, N., Lyytinen, H., Vaessen, A., & Blomer, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, 21. 4, 551–559.